

GenCore version 5.1.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 08:27:02 ; Search time 32.5 seconds  
(without alignments)  
49.200 Million cell updates/sec

Title: US-09-812-485A-42  
Perfect score: 66  
Sequence: 1 DNDISPFSGDQ 12

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002:\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	66	100.0	12	AAO20372	Peptide of a calci
2	66	100.0	15	AAU93724	Dental product bon
3	66	100.0	15	AAO20377	C-terminal amide
4	66	100.0	23	AAU93726	Dental product bon
5	66	100.0	23	AAO20379	C-terminal amide
6	66	100.0	40	AAU93703	Dental product bon
7	66	100.0	40	AAO20353	Protein of matrix
8	66	100.0	97	AAU93681	Dental product bon
9	66	100.0	97	AAO20331	Protein of matrix
10	66	100.0	430	AAV53812	Amino acid sequenc

11	66	100.0	430	22	AAAB62669	Truncated phosphat
12	66	100.0	509	22	AAAB82922	Human osteoregulin
13	66	100.0	525	22	AAAB82920	Human osteoregulin
14	66	100.0	525	22	AAAB62689	Phosphatonin polyp
15	66	100.0	525	23	ABBO8526	Human phosphatonin
16	66	100.0	525	23	AAE13227	Human oncogenic os
17	66	100.0	540	22	AAAB82923	Human osteoregulin
18	66	100.0	556	22	AAAB82921	Human osteoregulin
19	60	90.9	15	23	AAU93725	Dental product bon
20	60	90.9	15	23	AAO20378	C-terminal amide
21	60	90.9	33	23	AAU93710	Dental product bon
22	60	90.9	33	23	AAO20360	Peptide of matrix
23	60	90.9	435	22	AAAB82918	Rat osteoregulin.
24	60	90.9	441	22	AAAB82919	Mouse osteoregulin
25	55	83.3	19	23	AAE13230	Human OOM1 calcium
26	49	74.2	30	23	AAU93706	Dental product bon
27	49	74.2	30	23	AAO20356	Peptide of matrix
28	48.5	73.5	35	23	AAU93705	Dental product bon
29	48.5	73.5	35	23	AAO20355	Peptide of matrix
30	48.5	73.5	40	23	AAU93704	Dental product bon
31	48.5	73.5	40	23	AAO20354	Protein of matrix
32	44	66.7	275	22	AAAB93507	Human polypeptide,
33	41	62.1	94	22	ABG08384	Novel human diagno
34	41	62.1	2039	19	AAW56322	Haemophilus paraga
35	41	62.1	2042	19	AAW56319	Haemophilus paraga
36	40	60.6	594	22	ABG17844	Novel human diagno
37	40	60.6	961	22	ABG25046	Novel human diagno
38	39	59.1	15	23	AAU93723	Dental product bon
39	39	59.1	15	23	AAO20376	C-terminal amide
40	39	59.1	111	22	AAU72963	Mouse type II inte
41	39	59.1	300	20	AAU04930	Mycobacterium spec
42	39	59.1	347	20	AAU36366	Pseudomonas aerugi
43	39	59.1	577	20	AAU04931	Mycobacterium spec
44	38.5	58.3	793	22	ABB68973	Drosophila melanog
45	38	57.6	279	17	AAW00777	Thermitase G67S, G

ALIGNMENTS

RESULT 1  
AAO20372  
ID AAO20372 standard; peptide; 12 AA.  
XX AAO20372;  
XX  
XX 31-MAY-2002 (first entry)  
XX  
XX Peptide of a calcium binding motif #1.  
XX  
XX Bone growth; RGD motif; integrin binding motif; calcium binding motif;  
KW glycosaminoglycan binding motif; bone loss; renal phosphate excretion;  
KW alveolar; teeth; odontoblast; osteoclast; dental tissue; skeletal loss;  
KW weakness.  
XX  
XX Unidentified.  
XX  
XX WO200214360-A1.  
XX  
XX 21-FEB-2002.  
XX  
XX 14-AUG-2001; 2001WO-US255542.  
XX  
XX 16-AUG-2000; 2000US-0641034.  
XX 19-MAR-2001; 2001US-0812485.  
XX (BIGB-) BIG BEAR BIO INC.  
XX  
XX Kumagai Y, Blacher RW, Yoneda T;  
XX WPI; 2002-291971/33.  
XX  
XX New peptide compound useful for reducing bone loss, is capable of

CC	as well as the growth or recruiting of osteoblast or odontoblast cells
CC	on the surface of the new skeletal growth. AAU93681-AAU93726 represent
CC	bone growth enhancing peptides for use in dental products.
XX	
SQ	Sequence 15 AA;
	Query Match 100.0%; Score 66; DB 23; Length 15;
	Best Local Similarity 100.0%; Pred. No. 4.6e-05;
	Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 DNDISPFSGDQ 12
DB	4 DNDISPFSGDQ 15
RESULT 3	
AAO20377	AAO20377 standard; peptide; 15 AA.
XX	AAO20377;
XX	
DT	31-MAY-2002 (first entry)
XX	
DE	C-terminal amidated synthesised peptide D-00004.
KW	Bone growth; RGD motif; integrin binding motif; calcium binding motif;
KW	glycosaminoglycan binding motif; bone loss; renal phosphate excretion;
KW	alveolar; teeth; odontoblast; osteoclast; dental tissue; skeletal loss;
KW	weakness; D00004.
XX	
OS	Synthetic.
XX	
FH	Key Location/Qualifiers
FT	Modified-site 15
FT	/note= "C-terminal amide"
XX	
PN	WO200214360-A1.
XX	
PD	21-FEB-2002.
XX	
PX	14-AUG-2001; 2001WO-US25542.
PX	
PR	16-AUG-2000; 2000US-0641034.
PR	19-AUG-2001; 2001US-0812485.
XX	
PA	(BIGB-) BIG BEAR BIO INC.
XX	
PI	Kumagai Y, Blacher RW, Yoneda T;
XX	
DR	WPI; 2002-291971/33.
XX	
PT	New peptide compound useful for reducing bone loss, is capable of
PT	enhancing bone growth, and comprises an integrin binding motif,
PT	glycosaminoglycan binding motif or a calcium binding motif -
XX	
PS	Example 1; Page 23; 50pp; English.
XX	
CC	The invention relates to a peptide compound capable of enhancing bone
CC	growth, and comprising 10-50 amino acids in a sequence, where the amino
CC	acids are in D- or L- conformation and the sequence comprises a motif
CC	selected from an integrin binding motif, a glycosaminoglycan binding
CC	motif and a calcium binding motif. The peptide of the invention is useful
CC	for reducing bone loss and for reducing renal phosphate excretion in an
CC	individual. The peptide is useful for promoting regeneration of alveolar
CC	bone and/or teeth, and increases the number and activity of odontoblasts
CC	/osteoclasts that help form dental tissues. The peptide is also useful
CC	for treating or preventing a condition associated with skeletal loss or
CC	weakness. This sequence represents a C-terminal amidated synthesised
CC	peptide D-00004 of the invention.
XX	
SQ	Sequence 15 AA;
	Query Match 100.0%; Score 66; DB 23; Length 15;

Best Local Similarity 100.0%; Pred. No. 4.6e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DNDISPFSGDQ 12  
Db 4 DNDISPFSGDQ 15

RESULT 4  
AAU93726  
ID AAU93726 standard; peptide; 23 AA.  
XX AC AAU93726;  
XX 02-JUL-2002 (first entry)  
XX Dental product bone growth enhancing peptide #46.  
XX Dental product; toothpaste; mouthwash; dental floss; bone growth;  
KW integrin binding motif; RGD; skeletal disease; dental disease; tooth;  
KW alveolar bone growth; osteoblast; odontoblast; osteopathic.  
XX Synthetic.  
XX WO200213775-A1.  
XX PN 21-FEB-2002.  
XX 09-AUG-2001; 2001WO-US25101.  
XX PF 16-AUG-2000; 2000US-225879P.  
XX (BIGB-) BIG BEAR BIO INC.  
XX Yoneda T, Nomizu M, Kumagai Y;  
XX WPI; 2002-329525/36.  
XX Dental product useful for treating skeletal diseases e.g. dental  
PT diseases comprises a base material and a compound comprising specific  
PT amino acid in a sequence containing the integrin binding motif -  
XX  
XX Claim 7; Page 21; 44pp; English.  
XX The present invention relates to dental products such as toothpastes,  
CC mouthwash and dental floss comprising a base material and a compound  
CC which promotes bone growth. Such compounds are peptide sequences  
CC comprising 10-50 amino acids and containing an integrin binding  
CC motif such as RGD in the D- or L- form, preferably the L-configuration.  
CC The peptides of the invention are useful for treating or preventing  
CC skeletal diseases such as dental disease. The peptides enhance tooth  
CC and/or alveolar bone growth on areas where deterioration has occurred,  
CC as well as the growth or recruiting of osteoblast or odontoblast cells  
CC on the surface of the new skeletal growth. AAU93681-AAU93726 represent  
CC bone growth enhancing peptides for use in dental products.  
XX  
SQ Sequence 23 AA;  
Query Match 100.0%; Score 66; DB 23; Length 23;  
Best Local Similarity 100.0%; Pred. No. 7.4e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DNDISPFSGDQ 12  
Db 8 DNDISPFSGDQ 19

RESULT 5  
AAO20379  
ID AAO20379 standard; peptide; 23 AA.  
XX AC AAO20379;  
XX

DT 31-MAY-2002 (first entry)  
XX C-terminal amidated synthesised peptide D-00006.  
DE  
XX Bone growth; RGD motif; integrin binding motif; calcium binding motif;  
KW glycosaminoglycan binding motif; bone loss; renal phosphate excretion;  
KW alveolar; teeth; odontoblast; osteoclast; dental tissue; skeletal loss;  
KW weakness; D00006.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 23 /note= "C-terminal amide"  
FT  
XX  
XX WO200214360-A1.  
XX 21-FEB-2002.  
XX 14-AUG-2001; 2001WO-US25542.  
XX 16-AUG-2000; 2000US-0641034.  
PR 19-MAR-2001; 2001US-0812485.  
XX (BIGB-) BIG BEAR BIO INC.  
XX PA  
XX Kumagai Y, Blacher RW, Yoneda T;  
PI WPI; 2002-291971/33.  
XX  
XX New peptide compound useful for reducing bone loss, is capable of  
PT enhancing bone growth, and comprises an integrin binding motif,  
PT glycosaminoglycan binding motif or a calcium binding motif -  
XX  
XX Example 1; Page 15; 50pp; English.  
XX The invention relates to a peptide compound capable of enhancing bone  
CC growth, and comprising 10-50 amino acids in a sequence, where the amino  
CC acids are in D- or L- conformation and the sequence comprises a motif  
CC selected from an integrin binding motif, a glycosaminoglycan binding  
CC motif and a calcium binding motif. The peptide of the invention is useful  
CC for reducing bone loss and for reducing renal phosphate excretion in an  
CC individual. The peptide is useful for promoting regeneration of alveolar  
CC bone and/or teeth, and increases the number and activity of odontoblasts  
CC /osteoclasts that help form dental tissues. The peptide is also useful  
CC for treating or preventing a condition associated with skeletal loss or  
CC weakness. This sequence represents a C-terminal amidated synthesised  
CC peptide D-00006 of the invention.  
XX  
SQ Sequence 23 AA;  
Query Match 100.0%; Score 66; DB 23; Length 23;  
Best Local Similarity 100.0%; Pred. No. 7.4e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DNDISPFSGDQ 12  
Db 8 DNDISPFSGDQ 19

RESULT 6  
AAU93703  
ID AAU93703 standard; peptide; 40 AA.  
XX AC AAU93703;  
XX  
XX 02-JUL-2002 (first entry)  
XX Dental product bone growth enhancing peptide #23.  
DE Dental product; toothpaste; mouthwash; dental floss; bone growth;  
KW integrin binding motif; RGD; skeletal disease; dental disease; tooth;  
KW alveolar bone growth; osteoblast; odontoblast; osteopathic.

XX OS Synthetic.  
 XX PN WO200213775-A1.  
 XX XX 21-FEB-2002.  
 XX PD 09-AUG-2001; 2001WO-US25101.  
 XX PF 16-AUG-2000; 2000US-225879P.  
 XX PR (BIGB-) BIG BEAR BIO INC.  
 XX PA Yoneda T, Nomizu M, Kumagai Y;  
 XX PI WPI; 2002-329525/36.  
 XX DR Dental product useful for treating skeletal diseases e.g. dental  
 XX PT diseases comprises a base material and a compound comprising specific  
 XX PT amino acid in a sequence containing the integrin binding motif -  
 XX XX Claim 6; Page 20; 44pp; English.  
 XX XX The present invention relates to dental products such as toothpastes,  
 XX CC mouthwash and dental floss comprising a base material and a compound  
 XX CC which promotes bone growth. Such compounds are peptide sequences  
 XX CC comprising 10-50 amino acids and containing an integrin binding  
 XX CC motif such as RGD in the D- or L- form, preferably the L-configuration.  
 XX CC The peptides of the invention are useful for treating or preventing  
 XX CC skeletal diseases such as dental disease. The peptides enhance tooth  
 XX CC and/or alveolar bone growth on areas where deterioration has occurred,  
 XX CC as well as the growth or recruiting of osteoblast or odontoblast cells  
 XX CC on the surface of the new skeletal growth. AAU93681-AAU93726 represent  
 XX CC bone growth enhancing peptides for use in dental products.  
 XX SQ Sequence 40 AA;  
 Query Match 100.0%; Score 66; DB 23; Length 40;  
 Best Local Similarity 100.0%; Pred. No. 0.00014;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 DNDISPFSGDQG 12  
 Db 3 DNDISPFSGDQG 14  
 RESULT 7  
 AAU93681  
 ID AAU93681 standard; protein; 40 AA.  
 XX AC AAU93681;  
 XX AC AAU93681;  
 XX DT 31-MAY-2002 (first entry)  
 XX DE Protein of matrix extracellular phosphoglycoprotein containing RGD #23.  
 XX DE Bone growth; RGD motif; integrin binding motif; calcium binding motif;  
 XX KW glycosaminoglycan binding motif; bone loss; renal phosphate excretion;  
 XX KW alveolar; teeth; odontoblast; osteoclast; dental tissue; skeletal loss;  
 XX KW weakness; matrix extracellular phosphoglycoprotein.  
 XX OS Unidentified.  
 XX OS WO200214360-A1.  
 XX PN 21-FEB-2002.  
 XX PD 14-AUG-2001; 2001WO-US25542.  
 XX PF 16-AUG-2000; 2000US-0641034.  
 XX PR 19-MAR-2001; 2001US-0812485.  
 XX PR (BIGB-) BIG BEAR BIO INC.  
 XX PA

XX PI Kumagai Y, Blacher RW, Yoneda T;  
 XX DR WPI; 2002-291971/33.  
 XX XX New peptide compound useful for reducing bone loss, is capable of  
 XX PT enhancing bone growth, and comprises an integrin binding motif,  
 XX PT glycosaminoglycan binding motif or a calcium binding motif -  
 XX XX Disclosure; Page 12; 50pp; English.  
 XX XX The invention relates to a peptide compound capable of enhancing bone  
 XX CC growth, and comprising 10-50 amino acids in a sequence, where the amino  
 XX CC acids are in D- or L- conformation and the sequence comprises a motif  
 XX CC selected from an integrin binding motif, a glycosaminoglycan binding  
 XX CC motif and a calcium binding motif. The peptide of the invention is useful  
 XX CC for reducing bone loss and for reducing renal phosphate excretion in an  
 XX CC individual. The peptide is useful for promoting regeneration of alveolar  
 XX CC bone and/or teeth, and increases the number and activity of odontoblasts  
 XX CC /osteoclasts that help form dental tissues. The peptide is also useful  
 XX CC for treating or preventing a condition associated with skeletal loss or  
 XX CC weakness. This sequence represents a protein of a matrix extracellular  
 XX CC phosphoglycoprotein containing an RGD motif of the invention.  
 XX SQ Sequence 40 AA;  
 Query Match 100.0%; Score 66; DB 23; Length 40;  
 Best Local Similarity 100.0%; Pred. No. 0.00014;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 DNDISPFSGDQG 12  
 Db 3 DNDISPFSGDQG 14  
 RESULT 8  
 AAU93681  
 ID AAU93681 standard; protein; 97 AA.  
 XX AC AAU93681;  
 XX AC AAU93681;  
 XX DT 02-JUL-2002 (first entry)  
 XX DE Dental product bone growth enhancing peptide #1.  
 XX DE Dental product; toothpaste; mouthwash; dental floss; bone growth;  
 XX KW integrin binding motif; RGD; skeletal disease; dental disease; tooth;  
 XX KW alveolar bone growth; osteoblast; odontoblast; osteopathic.  
 XX OS Synthetic.  
 XX OS WO200213775-A1.  
 XX PN 21-FEB-2002.  
 XX PD 09-AUG-2001; 2001WO-US25101.  
 XX PF 16-AUG-2000; 2000US-225879P.  
 XX PR (BIGB-) BIG BEAR BIO INC.  
 XX PA Yoneda T, Nomizu M, Kumagai Y;  
 XX PI WPI; 2002-329525/36.  
 XX DR Dental product useful for treating skeletal diseases e.g. dental  
 XX PT diseases comprises a base material and a compound comprising specific  
 XX PT amino acid in a sequence containing the integrin binding motif -  
 XX XX Disclosure; Page 11; 44pp; English.  
 XX XX The present invention relates to dental products such as toothpastes,  
 XX CC mouthwash and dental floss comprising a base material and a compound



CC which promotes bone growth. Such compounds are peptide sequences  
 CC comprising 10-50 amino acids and containing an integrin binding  
 CC motif such as RGD in the D- or L- form. Preferably the L-configuration.  
 CC The peptides of the invention are useful for treating or preventing  
 CC skeletal diseases such as dental disease. The peptides enhance tooth  
 CC and/or alveolar bone growth on areas where deterioration has occurred,  
 CC as well as the growth or recruiting of osteoblast or odontoblast cells  
 CC on the surface of the new skeletal growth. AAU93681-AAU93726 represent  
 CC bone growth enhancing peptides for use in dental products.  
 XX  
 SQ Sequence 97 AA;

Query Match 100.0%; Score 66; DB 23; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 0.00039;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DNDISPFSGDQ 12  
 | | | | | | | | | |  
 Db 50 DNDISPFSGDQ 61

RESULT 9  
 AAO20331  
 ID AAO20331 standard; protein: 97 AA.

XX  
 AC AAO20331;

DT 31-MAY-2002 (first entry)

DE Protein of matrix extracellular phosphoglycoprotein containing RGD #1.  
 XX Bone growth; RGD motif; integrin binding motif; calcium binding motif;  
 KW glycosaminoglycan binding motif; bone loss; renal phosphate excretion;  
 KW alveolar; teeth; odontoblast; osteoclast; dental tissue; skeletal loss;  
 KW weakness; matrix extracellular phosphoglycoprotein.

XX Unidentified.

XX WO200214360-A1.

XX 21-FEB-2002.

XX 14-AUG-2001; 2001WO-US25542.

XX 16-AUG-2000; 2000US-0641034.

XX 19-MAR-2001; 2001US-0812485.

XX (BIGB-) BIG BEAR BIO INC.

XX Kumagai Y, Blacher RW, Yoneda T;

XX WPI; 2002-291971/33.

XX New peptide compound useful for reducing bone loss, is capable of  
 PT enhancing bone growth, and comprises an integrin binding motif,  
 PT glycosaminoglycan binding motif or a calcium binding motif -

PS Disclosure; Page 11; 50pp; English.

XX The invention relates to a peptide compound capable of enhancing bone  
 CC growth, and comprising 10-50 amino acids in a sequence, where the amino  
 CC acids are in D- or L- conformation and the sequence comprises a motif  
 CC selected from an integrin binding motif, a glycosaminoglycan binding  
 CC motif and a calcium binding motif. The peptide of the invention is useful  
 CC for reducing bone loss and for reducing renal phosphate excretion in an  
 CC individual. The peptide is useful for promoting regeneration of alveolar  
 CC bone and/or teeth, and increases the number and activity of odontoblasts  
 CC /osteoclasts that help form dental tissues. The peptide is also useful  
 CC for treating or preventing a condition associated with skeletal loss or  
 CC weakness. This sequence represents a protein of a matrix extracellular  
 CC phosphoglycoprotein containing an RGD motif of the invention.

XX Sequence 97 AA;

Query Match 100.0%; Score 66; DB 23; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 0.00039;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DNDISPFSGDQ 12  
 | | | | | | | | | |  
 Db 50 DNDISPFSGDQ 61

RESULT 10  
 AAY53812

ID AAY53812 standard; Protein: 430 AA.

XX  
 AC AAY53812;

DT 22-FEB-2000 (first entry)

XX Amino acid sequence of a human phosphatonin polypeptide.

XX Human; phosphatonin; Metastatic-tumour Excreted Phosphaturic-Element;  
 KW MEPE; Na+-dependent phosphate cotransport; vitamin D metabolism;  
 KW bone mineralisation; phosphate metabolism related disease;  
 KW hyperphosphatemia; renal osteodystrophy; renal dialysis;  
 KW secondary hyperparathyroidism; osteitis fibrosa cystica; hypercalcaemia;  
 KW X-linked hypophosphatemic ricket; hereditary hypophosphatemic ricket;  
 KW hypomineralised bone lesion; stunted growth; cystic fibrosis;  
 KW oncogenic hypophosphatemic osteomalacia; renal phosphate leakage;  
 KW renal osteodystrophy; osteoporosis; vitamin D resistant rickets;  
 KW end organ resistance; renal Fanconi syndrome; autosomal rickets;  
 KW Paget's disease; kidney failure; renal tubular acidosis; sprue.

XX Homo sapiens.

OS  
 FH Key Location/Qualifiers

FT Modified-site 8..10 /note= "protein kinase C phosphorylation site"  
 FT Modified-site 8..11 /note= "Casein kinase II phosphorylation site"  
 FT Modified-site 16..21 /note= "myristoylation site"  
 FT Modified-site 40..47 /note= "Tyrosine kinase phosphorylation site"  
 FT Modified-site 77..79 /note= "protein kinase C phosphorylation site"  
 FT Modified-site 118..120 /note= "protein kinase C phosphorylation site"  
 FT Modified-site 119..124 /note= "myristoylation site"  
 FT Modified-site 139..142 /note= "Casein kinase II phosphorylation site"  
 FT Modified-site 143..148 /note= "myristoylation site"  
 FT Region 152..154 /note= "cell attachment tripeptide"  
 FT Region 161..165 /note= "glycosaminoglycan attachment site"  
 FT Modified-site 177..180 /note= "Casein kinase II phosphorylation site"  
 FT Modified-site 194..197 /note= "Casein kinase II phosphorylation site"  
 FT Modified-site 199..202 /note= "Casein kinase II phosphorylation site"  
 FT Modified-site 203..205 /note= "protein kinase C phosphorylation site"  
 FT Modified-site 224..227 /note= "Casein kinase II phosphorylation site"  
 FT Modified-site 228..231 /note= "Casein kinase II phosphorylation site"  
 FT Modified-site 228..230 /note= "Casein kinase II phosphorylation site"  
 FT Modified-site 238..241 /note= "protein kinase C phosphorylation site"  
 FT Modified-site 238..241 /note= "Casein kinase II phosphorylation site"

FT Modified-site 266..271 /note="myristoylation site"  
 FT Modified-site 291..296 /note="myristoylation site"  
 FT Modified-site 311..313 /note="protein kinase C phosphorylation site"  
 FT Modified-site 312..314 /note="protein kinase C phosphorylation site"  
 FT Modified-site 315..320 /note="myristoylation site"  
 FT Modified-site 319..321 /note="protein kinase C phosphorylation site"  
 FT Modified-site 325..328 /note="Casein kinase II phosphorylation site"  
 FT Modified-site 370..373 /note="amidation site"  
 FT Modified-site 382..386 /note="Asu-glycosylation site"  
 FT Modified-site 383..387 /note="Asu-glycosylation site"  
 FT Modified-site 384..386 /note="protein kinase C phosphorylation site"  
 FT Modified-site 389..394 /note="myristoylation site"  
 FT Modified-site 403..405 /note="protein kinase C phosphorylation site"  
 FT Modified-site 405..408 /note="cAMP and cGMP dependent protein kinase phosphorylation site"  
 FT Modified-site 408..410 /note="protein kinase C phosphorylation site"  
 FT Modified-site 409..411 /note="protein kinase C phosphorylation site"  
 FT Modified-site 423..426 /note="Casein kinase II phosphorylation site"  
 FT Modified-site 425..428 /note="Casein kinase II phosphorylation site"  
 FT Modified-site 427..430 /note="Casein kinase II phosphorylation site"  
 XX WO9960017-A2.  
 PN 25-NOV-1999.  
 PD 18-MAY-1999; 99WO-EP03403.  
 PF 18-MAY-1998; 98GB-0010681.  
 PR 04-SEP-1998; 98GB-0019387.  
 XX (UNLO ) UNIV COLLEGE LONDON.  
 PA Rowe P;  
 XX WPI; 2000-053262/04.  
 DR N-PSDB; AAZ36447.  
 XX New polypeptides involved in the regulation of phosphate metabolism useful for diagnosing and treating disorders related to phosphate metabolism  
 PT Claim 6; Fig 8; 136pp; English.  
 PS The present sequence represents a phosphatonin polypeptide (also called Metastatic-tumour Excreted Phosphaturic-Element (MEPE)). The level of phosphatonin in a subject modulates Na+-dependent phosphate cotransport, vitamin D metabolism and/or bone mineralisation. The phosphatonin polypeptides, polynucleotides, vectors and antibodies are used to treat phosphate metabolism related disease. They are used for treatment of hyperphosphatemia, or renal osteodystrophy, hyperphosphatemia in renal dialysis/pre-dialysis, secondary hyperparathyroidism or osteitis fibrosa cystica, or X-linked hypophosphatemic rickets, hereditary hypophosphatemic rickets with hypocalcuria (HHRH), hypomineralised bone lesions, stunted growth in juveniles, oncogenic hypophosphatemic

CC osteomalacia, renal phosphate leakage, renal osteodystrophy, osteoporosis, vitamin D resistant rickets, end organ resistance, renal Fanconi syndrome, autosomal rickets, Paget's disease, kidney failure, renal tubular acidosis, cystic fibrosis or sprue. The polypeptide may also be used to manufacture combined preparations for simultaneous, separate or sequential use for the treatment of phosphate metabolism disorders. A transformed osteoblast or bone cell line capable of phosphatonin overexpression is useful for the production of phosphatonin.  
 XX Sequence 430 AA;  
 SQ Query Match 100.0%; Score 66; DB 21; Length 430;  
 Best Local Similarity 100.0%; Pred. No. 0.0021;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DNDISPFSGDQ 12  
 DB 154 DNDISPFSGDQ 165  
 RESULT 11  
 AAB62669  
 ID AAB62669 standard; Protein: 430 AA.  
 AC AAB62669;  
 XX 23-JUL-2001 (first entry)  
 DT Truncated phosphatonin polypeptide (truncated MEPE).  
 DE Metastatic-tumour excreted phosphaturic element; MEPE; phosphatonin;  
 KW phosphate; vitamin-D; skeletal formation; mineralization; truncated;  
 KW osteopathic; antigout; cytostatic; human.  
 XX Homo sapiens.  
 OS Key Location/Qualifiers  
 FH Misc-difference 161..192  
 FT /note="the amino acid residues in this region are indicated incorrectly in the sequence provided in the sequence listing"  
 FT WO200132878-A2.  
 PN 10-MAY-2001.  
 PD 31-OCT-2000; 2000WO-EP10747.  
 PF 04-NOV-1999; 99US-0434185.  
 PR 08-NOV-1999; 99GB-0026424.  
 XX (UNLO ) UNIV COLLEGE LONDON.  
 PA Rowe P;  
 XX WPI; 2001-343487/36.  
 DR N-PSDB; AAF83759.  
 XX New phosphatonin polypeptide a regulator of phosphate metabolism, for diagnosing and treating disorders of phosphate, vitamin-D metabolism, skeletal formation e.g. osteoporosis, Paget's disease, gout  
 PS Claim 4; Fig 8; 135pp; English.  
 XX The invention relates to a novel human protein, metastatic-tumour excreted phosphaturic element (MEPE) or phosphatonin (modulator of phosphate and vitamin-D metabolism). The phosphatonin (modulator of polynucleotides and specific antibodies are useful for treating a disorder of phosphate or vitamin D metabolism, skeletal formation and mineralization. Phosphatonins are used to treat hyperphosphatemia, renal osteodystrophy, secondary hyperparathyroidism, osteitis fibrosa cystica or gout. It is used to prepare a medicament for treating X-linked

CC hypophosphatemic rickets, hereditary hypophosphatemic rickets with  
 CC hypercalcuria (HHRH), hypomineralized bone lesions, stunted growth in  
 CC juveniles, oncogenic hypophosphatemic osteomalacia, renal phosphate  
 CC leakage, renal osteodystrophy, osteoporosis, vitamin-D resistant rickets,  
 CC end organ resistance, renal Fanconi syndrome, autosomal rickets, Paget's  
 CC disease, kidney failure, renal tubular acidosis, cystic fibrosis or  
 CC sprue. Phosphatonin polynucleotides are useful as molecular weight  
 CC markers on Southern gels, as diagnostic probes for detecting the presence  
 CC of a specific mRNA. Phosphatonin polypeptides are also useful for  
 CC identifying agonists and antagonists, compounds which bind to  
 CC phosphatonin and drug candidates for therapy of phosphate metabolism  
 CC disorders. The present sequence represents a truncated form of  
 CC phosphatonin (WEPE).  
 XX  
 XX

SQ Sequence 430 AA;  
 Query Match 100.0%; Score 66; DB 22; Length 430;  
 Best Local Similarity 100.0%; Pred. No. 0.0021;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DNDISPFSGDGO 12  
 ID 154 DNDISPFSGDGO 165  
 DB 154 DNDISPFSGDGO 165

RESULT 12  
 AAB82922  
 ID AAB82922 standard; Protein; 509 AA.

XX AAB82922;  
 XX 21-DEC-2001 (first entry)  
 XX Human osteoregulin (mature polypeptide).

XX Osteoregulin; human; bone; homeostasis; adipose; calcification;  
 KW atherosclerosis; osteoporosis; osteopathic; antiarteriosclerotic;  
 KW therapy.

XX Homo sapiens.

XX EP1130098-A2.

XX 05-SEP-2001.

XX 27-FEB-2001; 2001EP-0301768.

XX 29-FEB-2000; 2000US-185617P.

XX 22-SEP-2000; 2000US-234500P.

XX (PFIZ ) PFIZER PROD INC.

XX Brown TA, De Wet JR, Gowen LC, Hames LM;

XX WPI; 2001-604111/69.

XX N-PSDB; AAB26810.

XX Novel osteoregulin polypeptide useful for regulating bone homeostasis,  
 PT adiposity and calcification of atherosclerotic plaques comprises  
 PT measuring the activity of osteoregulin -

XX Claim 1; Page 54-55; 90pp; English.

XX The present sequence is that of human osteoregulin mature  
 CC polypeptide, i.e. lacking an N-terminal signal sequence.  
 CC Osteoregulin is a novel protein which plays a role in regulating  
 CC bone homeostasis, adiposity, and the calcification of  
 CC atherosclerotic plaques. 2 Splice variants of human osteoregulin  
 CC were identified (see also AAB82923). The invention provides novel  
 CC osteoregulin proteins, nucleic acids which encode them, vectors,  
 CC antibodies, host cells which express heterologous osteoregulins, and  
 CC animal cells and mammals with a targeted disruption of an  
 CC osteoregulin gene. The invention also provides screening assays

CC to identify modulators of osteoregulin activity as well as methods  
 CC of treating mammals for diseases or disorders associated with  
 CC osteoregulin activity. The modulators of activity may be useful  
 CC in the manufacture of a medicament for, as well as for treating, a  
 CC mammal in need of regulation of bone mass and/or density, adiposity, a  
 CC vascular flexibility, and/or atherosclerotic plaque calcification  
 CC (claimed), for treating and preventing osteoporosis, and for  
 CC stimulating bone repair and regeneration.  
 XX  
 XX

SQ Sequence 509 AA;

Query Match 100.0%; Score 66; DB 22; Length 509;  
 Best Local Similarity 100.0%; Pred. No. 0.0026;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DNDISPFSGDGO 12  
 ID 233 DNDISPFSGDGO 244  
 DB 233 DNDISPFSGDGO 244

RESULT 13  
 AAB82920  
 ID AAB82920 standard; Protein; 525 AA.

XX AAB82920;

XX 21-DEC-2001 (first entry)

XX Human osteoregulin.

XX Osteoregulin; human; bone; homeostasis; adipose; calcification;

KW atherosclerosis; osteoporosis; osteopathic; antiarteriosclerotic;  
 KW therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..37

FT /label= Signal\_peptide

FT Protein 38..525

FT /label= Mature\_protein

XX EP1130098-A2.

XX 05-SEP-2001.

XX 27-FEB-2001; 2001EP-0301768.

XX 29-FEB-2000; 2000US-185617P.

XX 22-SEP-2000; 2000US-234500P.

XX (PFIZ ) PFIZER PROD INC.

XX Brown TA, De Wet JR, Gowen LC, Hames LM;

XX WPI; 2001-604111/69.

XX N-PSDB; AAB26808.

XX Novel osteoregulin polypeptide useful for regulating bone homeostasis,  
 PT adiposity and calcification of atherosclerotic plaques comprises  
 PT measuring the activity of osteoregulin -

XX Claim 1; Page 45-47; 90pp; English.

XX The present sequence is that of human osteoregulin, a novel protein  
 CC which plays a role in regulating bone homeostasis, adiposity, and  
 CC the calcification of atherosclerotic plaques. The sequence is  
 CC predicted from the nucleotide sequence (see AAB26808) of isolated  
 CC osteoblast cDNA. A splice variant of human osteoregulin was also  
 CC identified (see AAB82921). The invention provides novel  
 CC osteoregulin proteins, nucleic acids which encode them, vectors,  
 CC antibodies, host cells which express heterologous osteoregulins, and  
 CC animal cells and mammals with a targeted disruption of an

CC osteoregulin gene. The invention also provides screening assays  
 CC to identify modulators of osteoregulin activity as well as methods  
 CC of treating mammals for diseases or disorders associated with  
 CC osteoregulin activity. The modulators of activity may be useful  
 CC in the manufacture of a medicament for, as well as for treating, a  
 CC mammal in need of regulation of bone mass and/or density, adiposity,  
 CC vascular flexibility, and/or atherosclerotic plaque calcification  
 CC (claimed), for treating and preventing osteoporosis, and for  
 CC stimulating bone repair and regeneration.  
 XX  
 SQ Sequence 525 AA;  
 Query Match 100.0%; Score 66; DB 22; Length 525;  
 Best Local Similarity 100.0%; Pred. No. 0.0027;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DNDISPFSGDQ 12  
 Db 249 DNDISPFSGDQ 260  
 RESULT 14  
 AAB62689  
 ID AAB62689 standard; Protein; 525 AA.  
 XX  
 AC AAB62689;  
 XX  
 DT 23-JUL-2001 (first entry)  
 DE Phosphatonin polypeptide (MEPE).  
 XX  
 KW Metastatic-tumour excreted phosphaturic element; MEPE; phosphatonin;  
 KW phosphate; vitamin-D; skeletal formation; mineralization; truncated;  
 KW osteopathic; antigout; cytostatic; human.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200132878-A2.  
 XX  
 PD 10-MAY-2001.  
 XX  
 PF 31-OCT-2000; 2000WO-EP10747.  
 XX  
 PR 04-NOV-1999; 99US-0434185.  
 PR 08-NOV-1999; 99GB-0026424.  
 XX  
 PA (UNLO ) UNIV COLLEGE LONDON.  
 XX  
 PI Rowe P;  
 XX  
 WPI; 2001-343487/36.  
 N-PSDB; AAF83764.  
 XX  
 PT New phosphatonin polypeptide a regulator of phosphate metabolism, for  
 PT diagnosing and treating disorders of phosphate, vitamin-D metabolism,  
 PT skeletal formation e.g. osteoporosis, Paget's disease, gout  
 XX  
 PS Claim 4; Page 133-134; 135pp; English.  
 XX  
 CC The invention relates to a novel human protein, metastatic-tumour  
 CC excreted phosphaturic element (MEPE) or phosphatonin (modulator of  
 CC phosphate and vitamin-D metabolism). The phosphatonin polypeptides,,  
 CC polynucleotides and specific antibodies are useful for treating a  
 CC disorder of phosphate or vitamin D metabolism, skeletal formation and  
 CC mineralization. Phosphatonins are used to treat hyperphosphatemia, renal  
 CC osteodystrophy, secondary hyperparathyroidism, osteitis fibrosa cystica  
 CC or gout. It is used to prepare a medicament for treating x-linked  
 CC hypophosphatemic rickets, hereditary hypophosphatemic rickets with  
 CC hypercalcaemia (HHRH), hypomineralized bone lesions, stunted growth in  
 CC juveniles, oncogenic hypophosphatemic osteomalacia, renal phosphate  
 CC leakage, renal osteodystrophy, osteoporosis, vitamin-D resistant rickets,  
 CC end organ resistance, renal Fanconi syndrome, autosomal rickets, Paget's  
 CC disease, kidney failure, renal tubular acidosis, cystic fibrosis or

CC sprue. Phosphatonin polynucleotides are useful as molecular weight  
 CC markers on Southern gels, as diagnostic probes for detecting the presence  
 CC of a specific mRNA. Phosphatonin polypeptides are also useful for  
 CC identifying agonists and antagonists, compounds which bind to  
 CC phosphatonin and drug candidates for therapy of phosphate metabolism  
 CC disorders. The present sequence represents the amino acid sequence of  
 CC the entire phosphatonin (MEPE).  
 XX  
 SQ Sequence 525 AA;  
 Query Match 100.0%; Score 66; DB 22; Length 525;  
 Best Local Similarity 100.0%; Pred. No. 0.0027;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DNDISPFSGDQ 12  
 Db 249 DNDISPFSGDQ 260  
 RESULT 15  
 ABB08526  
 ID ABB08526 standard; protein; 525 AA.  
 XX  
 AC ABB08526;  
 XX  
 DT 23-MAY-2002 (first entry)  
 DE Human phosphatonin.  
 XX  
 KW Human phosphatonin; cytostatic; antidiabetic; antiinflammatory;  
 KW hyperphosphatemia; arteriosclerosis; heart failure;  
 KW diabetic renal disease; kidney failure; cystic fibrosis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200198495-A1.  
 XX  
 PD 27-DEC-2001.  
 XX  
 PF 20-JUN-2001; 2001WO-JP05263.  
 XX  
 PR 21-JUN-2000; 2000JP-0191088.  
 XX  
 PA (TAKE ) TAKEDA CHEM IND LTD.  
 XX  
 PI Kurokawa T, Yamada T, Morimoto S;  
 XX  
 WPI; 2002-139791/18.  
 N-PSDB; ABA99159, ABA99160.  
 XX  
 PT Phosphatonin of human origin and DNA encoding it for diagnosis and  
 PT treatment of diseases associated with disorders of phosphate  
 PT metabolism, e.g., hyperphosphatemia, arteriosclerosis, heart failure,  
 PT diabetic renal disease and kidney failure  
 XX  
 PS Claim 1; Fig 1-4; 130pp; Japanese.  
 XX  
 CC This invention relates to human phosphatonin which has  
 CC the activity of lowering blood phosphate and increasing urinary  
 CC phosphate. The proteins are cytostatic, antidiabetic and  
 CC antiinflammatory in their action. Phosphatonin down-regulates  
 CC sodium-dependent phosphate transport in the kidney, it down-regulates  
 CC 25-hydroxy-vitamin D3-lalpa-hydroxylase in the kidney, it down-regulates  
 CC 25-hydroxy-vitamin D3-24-hydroxylase in the kidney. Phosphatonin is  
 CC useful in the diagnosis, treatment and prevention of phosphate  
 CC metabolism related diseases such as hyperphosphatemia, arteriosclerosis,  
 CC heart failure, diabetic renal disease, kidney failure, acute coronary  
 CC disease and cystic fibrosis. This sequence represents human  
 CC phosphatonin.  
 XX  
 SQ Sequence 525 AA;  
 Query Match 100.0%; Score 66; DB 23; Length 525;

Best Local Similarity 100.0%; Pred. No. 0.0027;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DNDISPFSGDGQ 12  
Db 249 DNDISPFSGDGQ 260  
|||||

Search completed: March 18, 2003, 08:29:49  
Job time : 33.5 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 08:29:30 ; Search time 11.5 Seconds  
(without alignments)  
30.702 Million cell updates/sec

Title: US-09-812-485A-42  
Perfect score: 66  
Sequence: 1 DNDISPFSGDGO 12

Scoring table: BIOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCRTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	56.1	256	5 PCT-US96-01314-55	Sequence 55, App1
2	37	56.1	394	1 US-07-637-870-1	Sequence 1, App1
3	37	56.1	394	1 US-07-637-399-1	Sequence 1, App1
4	37	56.1	394	1 US-07-640-476-5	Sequence 5, App1
5	37	56.1	394	1 US-08-112-703-1	Sequence 1, App1
6	37	56.1	394	6 5290690-2	Patent No. 5290690
7	37	56.1	799	1 US-08-054-077C-2	Sequence 2, App1
8	37	56.1	1198	4 US-09-199-637A-405	Sequence 405, App1
9	36	54.5	106	4 US-09-113-977C-42	Sequence 42, App1
10	36	54.5	106	4 US-09-113-977C-43	Sequence 43, App1
11	36	54.5	107	4 US-09-113-977C-41	Sequence 41, App1
12	36	54.5	1327	4 US-09-134-001C-3471	Sequence 3471, App1
13	36	54.5	441	5 PCT-US93-12588-98	Sequence 98, App1
14	36	54.5	441	5 PCT-US95-08071-98	Sequence 98, App1
15	36	54.5	553	4 US-08-997-251-4	Sequence 98, App1
16	36	54.5	555	2 US-08-453-702A-98	Sequence 98, App1
17	36	54.5	556	1 US-07-998-003A-98	Sequence 98, App1
18	36	54.5	556	1 US-08-453-374B-98	Sequence 98, App1
19	36	54.5	556	1 US-08-453-695A-98	Sequence 98, App1
20	36	54.5	556	1 US-08-268-161A-98	Sequence 98, App1
21	36	54.5	556	4 US-09-099-639-98	Sequence 98, App1
22	36	54.5	913	2 US-08-474-067-6	Sequence 6, App1
23	36	54.5	913	2 US-08-474-068A-6	Sequence 6, App1
24	36	54.5	913	2 US-08-472-481-5	Sequence 5, App1
25	36	54.5	1290	1 US-08-138-641-2	Sequence 2, App1
26	36	54.5	1290	1 US-08-138-133-2	Sequence 2, App1
27	36	54.5	1784	4 US-09-040-738-2	Sequence 2, App1

28	36	54.5	1784	4 US-08-652-426A-2	Sequence 2, App1
29	35	53.0	40	4 US-09-004-406C-16	Sequence 16, App1
30	35	53.0	137	4 US-09-036-574-4	Sequence 4, App1
31	35	53.0	211	2 US-08-708-958-1	Sequence 1, App1
32	35	53.0	240	4 US-09-355-166-19	Sequence 19, App1
33	35	53.0	278	2 US-08-865-203-6	Sequence 6, App1
34	35	53.0	278	2 US-07-849-420-6	Sequence 6, App1
35	35	53.0	278	4 US-09-233-854-6	Sequence 6, App1
36	35	53.0	278	4 US-08-955-424-6	Sequence 6, App1
37	35	53.0	279	2 US-09-135-658-4	Sequence 4, App1
38	35	53.0	279	6 5472855-6	Patent No. 5472855
39	35	53.0	532	4 US-09-294-841-2	Sequence 2, App1
40	35	53.0	604	4 US-09-391-104-30	Sequence 30, App1
41	35	53.0	607	4 US-09-000-041A-2	Sequence 2, App1
42	35	53.0	607	4 US-09-211-704A-10	Sequence 10, App1
43	35	53.0	1722	4 US-09-194-612A-1	Sequence 1, App1
44	34	51.5	106	4 US-09-113-977C-48	Sequence 48, App1
45	34	51.5	106	4 US-09-113-977C-49	Sequence 49, App1

#### ALIGNMENTS

```
RESULT 1
PCT-US96-01314-55
: Sequence 55, Application PC/TUS9601314
: GENERAL INFORMATION:
: APPLICANT: M. Amin Arnaout
: TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN
: NUMBER OF SEQUENCES: 78
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
: COMPUTER: IBM PS/2 Model 502 or 555X
: OPERATING SYSTEM: MS-DOS (Version 5.0)
: SOFTWARE: WordPerfect (Version 5.1)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US96/01314
: FILING DATE: 30-JAN-96
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/380,167
: FILING DATE: 30-JAN-95
: ATTORNEY/AGENT INFORMATION:
: NAME: John W. Freeman
: REGISTRATION NUMBER: 29,066
: REFERENCE/DOCKET NUMBER: 00786/267001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 542-5070
: TELEFAX: (617) 542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 55:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 256 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US96-01314-55

Query Match      56.1%; Score 37; DB 5; Length 256;
Best Local Similarity 87.5%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 DNDISPS 8
   | | | | | |
Db 64 DKDISPS 71
```

RESULT 2  
US-07-637-870-1  
Sequence 1, Application US/07637870  
Patent No. 5310665  
GENERAL INFORMATION:  
APPLICANT: Lambler, Anne-Marie  
APPLICANT: Lasters, Ignace  
APPLICANT: Quax, Wilhelms J.  
APPLICANT: Van Der Laan, Jan M.  
TITLE OF INVENTION: NOVEL GLUCOSE ISOMERASES HAVING  
TITLE OF INVENTION: ALTERED SUBSTRATE SPECIFICITY  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morrison & Foerster  
STREET: 545 Middlefield Road, Suite 200  
CITY: Menlo Park  
STATE: California  
COUNTRY: USA  
ZIP: 94025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/637,870  
FILING DATE: 19910104  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murashige, Kate H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 24615-20019.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-813-5600  
TELEFAX: 415-327-2951  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 394 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-637-870-1  
Query Match 56.1%; Score 37; DB 1; Length 394;  
Best Local Similarity 50.0%; Pred. No. 90;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
OY 1 DNDISPFSGDQ 12  
|:|:| | | |  
Db 55 DDLVFPFGSDAQ 66

RESULT 3  
US-07-637-399-1  
Sequence 1, Application US/07637399  
Patent No. 5340738  
GENERAL INFORMATION:  
APPLICANT: Lambelr, Anne-Marie  
APPLICANT: Lasters, Ignace  
APPLICANT: Mrabet, Nadir  
APPLICANT: Quax, Wilhelms J.  
APPLICANT: Van Der Laan, Jan M.  
TITLE OF INVENTION: NOVEL GLUCOSE ISOMERASES WITH AN  
TITLE OF INVENTION: ALTERED PH OPTIMUM  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morrison & Foerster  
STREET: 545 Middlefield Road, Suite 200  
CITY: Menlo Park

STATE: California  
COUNTRY: USA  
ZIP: 94025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/637,399  
FILING DATE: 19910104  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murashige, Kate H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 4615-0020.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-327-7250  
TELEFAX: 415-327-2951  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 394 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-637-399-1  
Query Match 56.1%; Score 37; DB 1; Length 394;  
Best Local Similarity 50.0%; Pred. No. 90;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
OY 1 DNDISPFSGDQ 12  
|:|:| | | |  
Db 55 DDLVFPFGSDAQ 66

RESULT 4  
US-07-640-476-5  
Sequence 5, Application US/07640476  
Patent No. 5376536  
GENERAL INFORMATION:  
APPLICANT: QUAX, WILHELMUS  
APPLICANT: LUTTEN, RUDOLF G.M.  
APPLICANT: SCHURHUIZEN, PAUL W.  
APPLICANT: MRABET, NADIR  
TITLE OF INVENTION: NOVEL GLUCOSE ISOMERASE ENZYMES AND  
TITLE OF INVENTION: THEIR USE  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morrison & Foerster  
STREET: 545 Middlefield Road, Suite 200  
CITY: Menlo Park  
STATE: CA  
COUNTRY: USA  
ZIP: 94025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/640,476  
FILING DATE: 19910110  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kate H. Murashige  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 24615-20009.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 327-7250  
TELEFAX: (415) 327-2951



TELEX: 706141  
; INFORMATION FOR SEQ. ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 394 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: 1linear  
; MOLECULE TYPE: protein  
US-07-640-476-5

Query Match 56.1%; Score 37; DB 1; Length 394;  
Best Local Similarity 50.0%; Pred. No. 90;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 DNDISPSGDAQ 12  
|:|:|1111|  
Db 55 DDDLVPFGSDAQ 66

RESULT 5  
US-08-112-703-1  
; Sequence 1, Application US/08112703  
; Patent No. 5384257  
; GENERAL INFORMATION:  
; APPLICANT: Lambell, Anne-Marie  
; APPLICANT: Lasters, Ignace  
; APPLICANT: Mrabet, Nadir  
; APPLICANT: Quax, Wilhelmus J.  
; APPLICANT: Van Der Laan, Jan M.  
; APPLICANT: Misset, Onno  
; TITLE OF INVENTION: NOVEL GLUCOSE ISOMERASES WITH AN  
; NUMBER OF SEQUENCES: 9  
; TITLE OF INVENTION: ALTERED PH OPTIMUM  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morrison & Foerster  
; STREET: 755 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/112,703  
; FILING DATE: 26-AUG-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murashige, Kate H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 4615002001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-813-5600  
; TELEFAX: 415-494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ. ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 394 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: 1linear  
US-08-112-703-1

Query Match 56.1%; Score 37; DB 1; Length 394;  
Best Local Similarity 50.0%; Pred. No. 90;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 DNDISPSGDAQ 12  
|:|:|1111|  
Db 55 DDDLVPFGSDAQ 66

RESULT 6  
5290690-2  
; Patent No. 5290690  
; APPLICANT: MRABET, NADIR; LASTERS, IGNACE; STANSSENS, PATRICK  
; MATTHYSSENS, GASTON; WODAK, SHOSHANA; QUAX, WILHELMUS J.  
; TITLE OF INVENTION: METHODS AND MEANS FOR CONTROLLING THE  
; STABILITY OF PROTEINS  
; NUMBER OF SEQUENCES: 22  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/398,706  
; FILING DATE: 25-AUG-1989  
; SEQ ID NO: 2:  
; LENGTH: 394  
5290690-2

Query Match 56.1%; Score 37; DB 6; Length 394;  
Best Local Similarity 50.0%; Pred. No. 90;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 DNDISPSGDAQ 12  
|:|:|1111|  
Db 55 DDDLVPFGSDAQ 66

RESULT 7  
US-08-054-077C-2  
; Sequence 2, Application US/08054077C  
; Patent No. 5527679  
; GENERAL INFORMATION:  
; APPLICANT: HEMLER, MARTIN E.  
; APPLICANT: RAMASWAMY, HEMAVATHI  
; TITLE OF INVENTION: HUMAN INTEGRIN 5 SUBUNIT PROTEIN  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &  
; STREET: 130 WATER STREET  
; CITY: BOSTON  
; STATE: MASSACHUSETTS  
; COUNTRY: US  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/054,077C  
; FILING DATE: 27-APR-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/694314  
; FILING DATE: 01-MAY-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RESNICK, DAVID S.  
; REGISTRATION NUMBER: 34235  
; REFERENCE/DOCKET NUMBER: 40937  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 523-3400  
; TELEFAX: (617) 523-6440  
; TELEX: 200291 STRE UR  
; INFORMATION FOR SEQ. ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 799 amino acids  
; TYPE: amino acid  
; TOPOLOGY: 1linear  
; MOLECULE TYPE: protein  
US-08-054-077C-2

Query Match 56.1%; Score 37; DB 1; Length 799;  
Best Local Similarity 87.5%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 DNDISPF8  
| | | | |  
Db 184 DKDISPFS 191

## RESULT 8

US-09-199-637A-405  
; Sequence 405, Application US/09199637A  
; Patent No. 6355411  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick  
; APPLICANT: Goodman, Howard M.  
; APPLICANT: Rahme, Laurence G.  
; APPLICANT: Mahajan-Miklos, Shalina  
; APPLICANT: Tan, Man-Wah  
; APPLICANT: Cao, Hui  
; APPLICANT: Drenkard, Eliana  
; APPLICANT: Tsongalis, John  
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
; FILE REFERENCE: 00786/361002  
; CURRENT APPLICATION NUMBER: US/09/199,637A  
; CURRENT FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,517  
; PRIOR FILING DATE: 1997-11-25  
; NUMBER OF SEQ ID NOS: 437  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 405  
; LENGTH: 1198  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-199-637A-405

Query Match 56.1%; Score 37; DB 4; Length 1198;  
Best Local Similarity 57.1%; Pred. No. 3e+02;  
Matches 8; Conservative 1; Mismatches 1; Indels 4; Gaps 1;

OY 1 DNDISPF---SGD 10  
| | | | |  
Db 407 DNDLKFPMYKSGD 420

## RESULT 9

US-09-113-977C-42  
; Sequence 42, Application US/09113977C  
; Patent No. 6277824  
; GENERAL INFORMATION:  
; APPLICANT: Doherty, Patrick  
; APPLICANT: Blaschuk, Orest W.  
; APPLICANT: Gour, Barbara J.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING ADHESION MOLECULE  
; FILE REFERENCE: 100086,403  
; CURRENT APPLICATION NUMBER: US/09/113,977C  
; CURRENT FILING DATE: 1998-07-10  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 42  
; LENGTH: 106  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-113-977C-42

Query Match 54.5%; Score 36; DB 4; Length 106;  
Best Local Similarity 63.6%; Pred. No. 32;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 DNDISPFSGDG 11  
| | | | |  
Db 82 DNGIPPMSTGTG 92

## RESULT 10

US-09-113-977C-43  
; Sequence 43, Application US/09113977C  
; Patent No. 6277824  
; GENERAL INFORMATION:  
; APPLICANT: Doherty, Patrick  
; APPLICANT: Blaschuk, Orest W.  
; APPLICANT: Gour, Barbara J.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING ADHESION MOLECULE  
; FILE REFERENCE: 100086,403  
; CURRENT APPLICATION NUMBER: US/09/113,977C  
; CURRENT FILING DATE: 1998-07-10  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 43  
; LENGTH: 106  
; TYPE: PRT  
; ORGANISM: Bos taurus  
US-09-113-977C-43

Query Match 54.5%; Score 36; DB 4; Length 106;  
Best Local Similarity 63.6%; Pred. No. 32;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 DNDISPFSGDG 11  
| | | | |  
Db 82 DNGIPPMSTGTG 92

## RESULT 11

US-09-113-977C-41  
; Sequence 41, Application US/09113977C  
; Patent No. 6277824  
; GENERAL INFORMATION:  
; APPLICANT: Doherty, Patrick  
; APPLICANT: Blaschuk, Orest W.  
; APPLICANT: Gour, Barbara J.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING ADHESION MOLECULE  
; FILE REFERENCE: 100086,403  
; CURRENT APPLICATION NUMBER: US/09/113,977C  
; CURRENT FILING DATE: 1998-07-10  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 41  
; LENGTH: 107  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-113-977C-41

Query Match 54.5%; Score 36; DB 4; Length 107;  
Best Local Similarity 63.6%; Pred. No. 32;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 DNDISPFSGDG 11  
| | | | |  
Db 83 DNGIPPMSTGTG 93

## RESULT 12

US-09-134-001C-3471  
; Sequence 3471, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779

;; PRIOR FILING DATE: 1997-08-14  
;; NUMBER OF SEQ ID NOS: 5674  
;; SEQ ID NO 3471  
;; LENGTH: 327  
;; TYPE: PRT  
;; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3471

Query Match 54.5%; Score 36; DB 4; Length 327;  
Best Local Similarity 77.8%; Pred. No. 1.5e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 DNDISPFSG 9  
||| ||| ||  
Db 199 DNDISPCSG 207

RESULT 13  
PCT-US93-12588-98  
; Sequence 98, Application PC/TUS9312588  
; GENERAL INFORMATION:  
; APPLICANT: Suzuki, Shintaro  
; TITLE OF INVENTION: Protocadherin Materials and Methods  
; NUMBER OF SEQUENCES: 107  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/12588  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/998,003  
; FILING DATE: 29 DEC 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Noland, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 31811  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 98:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 441 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
PCT-US93-12588-98

Query Match 54.5%; Score 36; DB 5; Length 441;  
Best Local Similarity 63.6%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 DNDISPFSGDG 11  
||| ||| |||  
Db 305 DNGIPMSGTG 315

RESULT 14  
PCT-US95-08071-98  
; Sequence 98, Application PC/TUS9508071

;; GENERAL INFORMATION:  
;; APPLICANT: Suzuki, Shintaro  
;; TITLE OF INVENTION: Protocadherin Materials and Methods  
;; NUMBER OF SEQUENCES: 115  
;; CORRESPONDENCE ADDRESSES:  
;; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
;; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
;; CITY: Chicago  
;; STATE: Illinois  
;; COUNTRY: USA  
;; ZIP: 60606  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US95/08071  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US93/12588  
;; FILING DATE: 23 DEC 1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/998,003  
;; FILING DATE: 29 DEC 1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Noland, Greta E.  
;; REGISTRATION NUMBER: 35,302  
;; REFERENCE/DOCKET NUMBER: 32149  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 312/474-6300  
;; TELEFAX: 312/474-0448  
;; TELEX: 25-3856  
;; INFORMATION FOR SEQ ID NO: 98:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 441 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: CDNA  
PCT-US95-08071-98

Query Match 54.5%; Score 36; DB 5; Length 441;  
Best Local Similarity 63.6%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 DNDISPFSGDG 11  
||| ||| |||  
Db 305 DNGIPMSGTG 315

RESULT 15  
US-08-997-251-4  
; Sequence 4, Application US/08997251  
; Patent No. 6271440  
; GENERAL INFORMATION:  
; APPLICANT: GUBLER, FRANZ J.  
; ATTORNEY/AGENT INFORMATION:  
; NAME: JACOBSEN, JOHN V.  
; TITLE OF INVENTION: PLANT REGULATORY PROTEINS III  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
; STREET: 5370 Manhattan Boulevard  
; CITY: Boulder  
; STATE: CO  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/997,251  
FILING DATE: 23-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO AU96/00383  
FILING DATE: 21-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PN6470/95  
FILING DATE: 09-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PN3779/95  
FILING DATE: 23-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Wanner, Ellen P.  
REGISTRATION NUMBER: 28,547  
REFERENCE/DOCKET NUMBER: 110-97  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ. ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 553 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-997-251-4

Query Match 54.5%; Score 36; DB 4; Length 553;  
Best Local Similarity 66.7%; Pred. No. 1.9e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 NDISPFGD 10  
||:||||:  
DB 432 NDCAPFSGN 440

Search completed: March 18, 2003, 08:32:21  
Job time: 13.5 secs

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 08:28:44 ; Search time 13.5 seconds  
(without alignments)  
85.453 Million cell updates/sec

Title: US-09-812-485a-42  
Perfect score: 66  
Sequence: 1 NNDISFSGDQ 12

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	68.2	913	2 D82885	multiple banded an
2	43	65.2	629	2 T39285	probable transmem
3	42	63.6	420	2 AE1857	hypothetical prote
4	41	62.1	337	2 S38166	hypothetical prote
5	41	62.1	886	2 G65036	hypothetical prote
6	40	60.6	1758	2 F88559	protein C48B4.4b l
7	40	60.6	1767	2 S60124	transport protein
8	39	59.1	300	2 H70879	dihydridipicolinat
9	39	59.1	347	2 A83170	S-adenosylmethioni
10	39	59.1	393	2 S59499	cellulase egli - s
11	39	59.1	2733	2 S15760	genome polypotein
12	38	57.6	348	2 C91060	hypothetical prote
13	38	57.6	444	2 B83891	intracellular alka
14	38	57.6	886	2 A85905	hypothetical prote
15	38	57.6	886	2 AD0831	probable acyl-CoA
16	37	56.1	394	1 ISMAXM	xylose isomerase (
17	37	56.1	402	1 F48552	protein kinase (EC
18	37	56.1	490	2 E69066	conserved hypothet
19	37	56.1	497	1 S20174	pectate lyase (EC
20	37	56.1	542	2 T06728	gene pointed prote
21	37	56.1	623	1 S33167	integrin beta-5 ch
22	37	56.1	656	2 JC3005	flagellar hook-ass
23	37	56.1	683	2 D83511	hypothetical prote
24	37	56.1	696	2 S44912	probable serine/ch
25	37	56.1	734	2 S37998	integrin beta-5 ch
26	37	56.1	799	2 A38308	T518.5 protein - A
27	37	56.1	832	2 C86431	DNA-directed RNA p
28	37	56.1	919	1 RNYZCA	yemaneuclein-alpha
29	37	56.1	1002	2 A56678	

30	37	56.1	2731	1 VFTHJH	genome polypotein
31	36	54.5	173	1 GOGAAA	atrial gland pepti
32	36	54.5	173	1 B45932	atrial gland pepti
33	36	54.5	173	1 GOGABA	hypothetical prote
34	36	54.5	237	2 B84377	hypothetical prote
35	36	54.5	358	2 C95898	hypothetical prote
36	36	54.5	378	2 C97611	hypothetical prote
37	36	54.5	378	2 AG2833	hypothetical prote
38	36	54.5	437	2 S11497	branched-chain ami
39	36	54.5	464	2 B35159	arylsulfatase (EC
40	36	54.5	464	2 H90140	transcription regu
41	36	54.5	507	2 G89908	Catalase (imported
42	36	54.5	553	2 T03762	myb-related transc
43	36	54.5	577	2 T45548	arylsulfatase (EC
44	36	54.5	594	2 F82895	hypothetical prote
45	36	54.5	623	2 G95180	ABC transporter, A

## ALIGNMENTS

```

RESULT 1
D82885      multiple banded antigen homolog U0487 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: D82885
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Helner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a
A:Reference number: A82870
A:Accession: D82885
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-913 <GLA>
A:Cross-references: GB:AE002146; GB:AF222894; NID:g6899479; PIDN:AAF30899.1; GSPDB:GN
C:Genetics:
A:Experimental source: serovar 3; biovar 1
A:Gene: U0487
A:Genetic code: SGC3

Query Match      68.2% Score 45; DB 2; Length 913;
Best Local Similarity 81.8%; Pred. No. 9.1;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 NDISFSGDQ 12
      ||||| |
DB      891 NDISPFSKDDQ 901

RESULT 2
T39285      probable transmembrane protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
C:Accession: T39285
R:McDougal, R.C.; Rajandream, M.A.; Barrell, B.G.; Ransperger, U.; Bothe, G.; Pohl,
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21841
A:Accession: T39285
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-629 <MCD>
A:Cross-references: EMBL:AL096851; PIDN:CAB50971.1; GSPDB:GN00067; SPDB:SPBCL105.08
A:Experimental source: strain 972h-; cosmid c1105
C:Genetics:
A:Gene: SPDB:SPBCL105.08
A:Map position: 2
A:Introns: 67/1
C:Superfamily: Schizosaccharomyces pombe probable transmembrane protein SPBCL105.08

Query Match      65.2% Score 43; DB 2; Length 629;
Best Local Similarity 72.7%; Pred. No. 14;

```

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 NDISPFGSDGQ 12  
 | | | | | | | |  
 Db 48 NTISPFIDGR 58

## RESULT 3

AE1857

hypothetical protein all0406 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

C:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C&gt;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C:Accession: AE1857

R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AE1857

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-420 &lt;BLAT&gt;

A:Cross-references: GB:BA000019; PIDN:BA072364.1; PID:g17129751; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all0406

Query Match 63.6%; Score 42; DB 2; Length 420;  
 Best Local Similarity 70.0%; Pred. No. 13;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 NDISPFGSDG 11  
 | | | | | | | |  
 Db 60 NGINPAGDG 69

## RESULT 4

S38166

hypothetical protein YKR088c - Yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YKR408

C:Species: Saccharomyces cerevisiae

C&gt;Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 19-Apr-2002

C:Accession: S38166; S42017; S39129

R:Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo, J.

submitted to the Protein Sequence Database, March 1994

A:Reference number: S38158

A:Accession: S38166

A:Molecule type: DNA

A:Residues: 1-337 &lt;BAL&gt;

A:Cross-references: EMBL:Z28313; NID:g486572; PIDN:CAA82167.1; PID:g486573; MIPS:YKR088c

A:Experimental source: strain S288C

R:Garcia-Cantalejo, J.; Baladron, V.; Esteban, P.F.; Santos, M.A.; Bou, G.; Remacha, M.A.

Yeast 10, 231-245, 1994

A:Title: The complete sequence of an 18,002 bp segment of Saccharomyces cerevisiae chrom

A:Reference number: S42009; MUID:94262327; PMID:8203164

A:Accession: S42017

A:Molecule type: DNA

A:Residues: 1-337 &lt;GAR&gt;

A:Cross-references: EMBL:Z27116; NID:g415899; PIDN:CAA81639.1; PID:g415908

A:Experimental source: strain S288C

C:Genetics:

A:Cross-references: SGD:S0001796

A:Map position: 11R

C:Keywords: transmembrane protein

F:96-112/Domain: transmembrane #status predicted &lt;TM1&gt;

F:138-154/Domain: transmembrane #status predicted &lt;TM2&gt;

F:174-190/Domain: transmembrane #status predicted &lt;TM3&gt;

F:290-306/Domain: transmembrane #status predicted &lt;TM4&gt;

Query Match 62.1%; Score 41; DB 2; Length 337;  
 Best Local Similarity 66.7%; Pred. No. 16;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 NDISPFGSDGQ 12  
 | | | | | | | |  
 Db 28 DNDIMPNNSNGQ 39

## RESULT 5

G65036

hypothetical protein b2584 - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C&gt;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C:Accession: G65036

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.: Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G65036

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-886 &lt;BLAT&gt;

A:Cross-references: GB:AE000344; GB:U00096; NID:g1788927; PIDN:AA075637.1; PID:g17889

A:Experimental source: strain K-12, substrain MG1655

Query Match 62.1%; Score 41; DB 2; Length 886;  
 Best Local Similarity 70.0%; Pred. No. 45;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 DISPFGSDGQ 12  
 | | | | | | | |  
 Db 693 DISPFGDNE 702

## RESULT 6

F88559

protein C48B4.4b [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C&gt;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 17-May-2002

C:Accession: F88559

R:anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio

A:Reference number: A75000; MUID:99065613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C\_

A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;

A:Accession: F88559

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1758 &lt;STO&gt;

A:Cross-references: GB:chr\_III; PIDN:CAA82384.1; PID:g3875025; GSPDB:GN00021; CESP:C4

C:Genetics:

A:Gene: C48B4.4b

A:Map position: 3

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog

Query Match 60.6%; Score 40; DB 2; Length 1758;  
 Best Local Similarity 63.6%; Pred. No. 1,4e+02;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 NDISPFGSDGQ 12  
 | | | | | | | |  
 Db 46 NDSTPLRGDQ 56

## RESULT 7

S60124

transport protein homolog C48B4.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C&gt;Date: 13-Jan-1996 #sequence\_revision 12-Apr-1996 #text\_change 02-Feb-2001

C:Accession: S60124; S40724; S40725

R:Kershaw, J.

submitted to the EMBL data library, November 1995

A:Reference number: S60124

A:Accession: S60124  
A:Molecule type: DNA  
A:Residues: 1-1767 <KER>  
A:Cross-references: EMBL:Z29117; NID:9439247; PID:q1066912  
C:Genetics:  
A:Map position: III  
A:Introns: 47/1; 112/3; 161/2; 220/2; 543/3; 574/3; 903/2; 1056/1; 1115/3; 1178/3; 1265/4  
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology  
C:Keywords: ATP; duplication; nucleotide binding; P-loop; transmembrane protein  
F:628-818/Domain: ATP-binding cassette homology <ABCL>  
F:645-652/Region: nucleotide-binding motif A (P-loop)  
F:764-769/Region: nucleotide-binding motif B  
F:1457-1642/Domain: ATP-binding cassette homology <ABC2>  
F:1474-1481/Region: nucleotide-binding motif A (P-loop)  
F:1586-1591/Region: nucleotide-binding motif B

Query Match 60.6%; Score 40; DB 2; Length 1767;  
Best Local Similarity 63.6%; Pred. No. 1.4e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 NDISPFGSGDQ 12  
||:|||||  
Db 46 NDSTPLRGDQ 56

RESULT 8  
H70879  
dihydropicolinate synthase (EC 4.2.1.52) RV2753c [similarity] - Mycobacterium tuberculosis  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: H70879  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajadurai, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.; Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: H70879  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-300 <COL>  
A:Cross-references: GB:AL008967; GB:AL123456; NID:93261491; PIDN:CAA15549.1; PID:9262427  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: dapa  
C:Function:  
A:Description: catalyzes the formation of dihydropicolinate from L-aspartate 4-semiald  
A:Pathway: lysine biosynthesis  
A:Note: first step in diaminopelate and lysine biosynthesis; sensitive to lysine inhib  
C:Superfamily: dihydropicolinate synthase  
C:Keywords: carbon-oxygen lyase; hydro-lyase; lysine biosynthesis  
F:171/Active site: Lys (covalent pyruvate-binding) #status predicted

Query Match 59.1%; Score 39; DB 2; Length 300;  
Best Local Similarity 75.0%; Pred. No. 31;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 4 ISPFSGDG 11  
:::|||||  
Db 20 VTFPSGDG 27

RESULT 9  
AB3170  
S-adenosylmethionine-tRNA ribosyltransferase-isomerase (EC 5.4.99.-) quoa PA3824 [simila  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: AB3170  
R:Stover, C.K.; Plam, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic p  
A:Reference number: AB2950; MUID:20437337; PMID:10984043  
A:Accession: AB3170  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-347 <STO>  
A:Cross-references: GB:AE004799; GB:AE004091; NID:99499981; PIDN:AG07211.1; GSPDB:G  
C:Experimental source: strain PA01  
C:Genetics:  
A:Gene: quoa; PA3824  
C:Superfamily: Escherichia coli tRNA ribosyltransferase-isomerase  
C:Keywords: intramolecular transferase; isomerase

Query Match 59.1%; Score 39; DB 2; Length 347;  
Best Local Similarity 60.0%; Pred. No. 37;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 DNDISPFSGD 10  
|::|||||  
Db 263 DGEIKPFSGD 272

RESULT 10  
S59499  
cellulase egII - smut fungus (Ustilago maydis)  
C:Species: Ustilago maydis (corn smut)  
C:Date: 20-Jul-1996 #sequence\_revision 13-Mar-1997 #text\_change 28-May-1999  
C:Accession: S59499  
R:Schauwecker, F.; Wanner, G.; Kahmann, R.  
Biol. Chem. Hoppe-Seyler 376, 617-625, 1995  
A:Title: Filament-specific expression of a cellulase gene in the dimorphic fungus U  
A:Reference number: S59499; MUID:96145728; PMID:8590631  
A:Accession: S59499  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-393 <SCH>  
A:Cross-references: GB:S81598; NID:91478378; PIDN:AAB36147.1; PID:91478379

Query Match 59.1%; Score 39; DB 2; Length 393;  
Best Local Similarity 77.8%; Pred. No. 42;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 DNDISPFSG 9  
||:|||||  
Db 239 DNTISPSYG 247

RESULT 11  
S15760  
genome polyprotein - murine hepatitis virus (strain A59)  
N:Alternate names: protein 1b  
N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)  
C:Species: murine hepatitis virus, MHV  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 20-Jun-2000  
C:Accession: S15760; S08652  
R:Bredius, P.J.; Pachuk, C.J.; Noten, A.F.H.; Charite, J.; Luytjes, W.; Weiss, S.  
Nucleic Acids Res. 18, 1825-1832, 1990  
A:Title: The primary structure and expression of the second open reading frame of th  
frameshifting mechanism.  
A:Reference number: S15760; MUID:90245573; PMID:2159623  
A:Accession: S15760  
A:Molecule type: genomic RNA  
A:Residues: 1-2733 <BRE>  
A:Cross-references: EMBL:X51939; NID:958974; PID:q1334829  
C:Superfamily: Infectious bronchitis virus RNA-directed RNA polymerase  
C:Keywords: nucleotidyltransferase; RNA biosynthesis

Query Match 59.1%; Score 39; DB 2; Length 2733;  
Best Local Similarity 60.0%; Pred. No. 3.4e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 DNDISPFSGD 10  
||:|||||





GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 08:27:42 ; Search time 7.5 Seconds  
(without alignments)  
66.362 Million cell updates/sec

Title: US-09-812-485a-42  
Perfect score: 66  
Sequence: 1 DNDISPFSGDQ 12

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	62.1	337	YK68_YEAST	P36164 saccharomyc
2	41	62.1	886	YF10_ECOLI	P76594 escherichia
3	39	59.1	300	DAPA_MYCTO	O33295 mycobacteri
4	39	59.1	347	QUPA_PSEAE	O9hxb8 pseudomonas
5	39	59.1	393	GUNI_USTMA	P54424 ustilago ma
6	39	59.1	2733	RREP_CWMA5	P16342 murine coro
7	37	56.1	329	IPNS_STRJU	P18286 streptomyce
8	37	56.1	393	XYLA_ACTMI	P12851 actinoplane
9	37	56.1	402	U120_HSMVG	O05101 marek's dis
10	37	56.1	497	MEK1_YEAST	P24719 saccharomyc
11	37	56.1	623	PNT1_DROME	P51022 drosophila
12	37	56.1	655	ITB5_PAPCY	O07441 papio cynoc
13	37	56.1	696	Y012_CAEEL	P34668 caenorhabdi
14	37	56.1	734	KKO8_YEAST	P36004 saccharomyc
15	37	56.1	798	ITB5_MOUSE	O70309 mus musculu
16	37	56.1	799	ITB5_HUMAN	P18084 homo sapien
17	37	56.1	919	RPO2_CAPVK	P16716 capripoxvir
18	37	56.1	1002	YEMA_DROME	P25992 drosophila
19	37	56.1	2731	RREP_CWMAH	P29982 murine coro
20	36	54.5	54	ELHB_APLCA	P11924 aplysia cal
21	36	54.5	54	ELHC_APLCA	P11925 aplysia cal
22	36	54.5	173	ELNA_APLCA	P01360 aplysia cal
23	36	54.5	209	COEC_SCYCA	P81130 scyllorhinu
24	36	54.5	263	PSBP_CUCSA	O9s1q8 cucumis sat
25	36	54.5	378	MURC_AGRT5	O8udm8 agrobacteri
26	36	54.5	437	BRAB_PSEAE	P19072 pseudomonas
27	36	54.5	464	ARS_KLEAE	P20713 klebsiella
28	36	54.5	591	AGPL_ASCSU	O17045 ascaris suu
29	36	54.5	823	SYL_DETRA	O9sf60 denococcus
30	36	54.5	877	CAD2_BOVIN	P19534 bos taurus
31	36	54.5	906	CAD2_HUMAN	P19022 homo sapien
32	36	54.5	906	CAD2_MOUSE	P15116 mus musculu
33	36	54.5	906	CAD2_RAT	O9z1y3 rattus norv

34	36	54.5	912	1	CAD2-CHICK	P10288 gallus gall
35	36	54.5	916	1	TOP1_ARATH	P30181 arabidopsis
36	36	54.5	1290	1	PIG1_HUMAN	P19174 homo sapien
37	36	54.5	1290	1	PIG1_RAT	P10686 rattus norv
38	36	54.5	1581	1	TIAM_HUMAN	O13009 homo sapien
39	36	54.5	1807	1	TSC2_HUMAN	P49815 homo sapien
40	36	54.5	1809	1	TSC2_RAT	P49816 rattus norv
41	36	54.5	1814	1	TSC2_MOUSE	O61037 mus musculu
42	35	53.0	179	1	FLIN_AGRT5	O57259 agrobacteri
43	35	53.0	205	1	KGRN_CHLTR	O84033 chlamydia t
44	35	53.0	271	1	SDRI_PICAB	O08632 picea abies
45	35	53.0	279	1	THER_THERVU	P04072 thermoactin

## ALIGNMENTS

RESULT 1	ID	YK68_YEAST	STANDARD:	PRT:	337 AA.
AC	P36164:				
DT	01-JUN-1994 (Rel. 29, Created)				
DT	01-JUN-1994 (Rel. 29, Last sequence update)				
DE	Hypothetical 38.3 kDa protein in PRP16-SRP40 intergenic region.				
GN	YKR088C OR YKR408.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OX	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
NC	NCBI_TaxID=4932;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=94262327; PubMed=8203164;				
RA	Garcia-Cantalejo J., Baladron V., Esteban P.F., Santos M.A., Bou G.,				
RA	Remacha M.A., Revuelta J.L., Ballesta J.P.G., Jimenez A., del Rey F.;				
RT	"The complete sequence of an 18,002 bp segment of Saccharomyces				
RT	cerevisiae chromosome XI contains the HBS1, MRP-L20 and PRP16 genes,				
RT	and six new open reading frames.";				
RL	Yeast 10:231-245(1994).				
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).				
CC					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC					
DR	EMBL: Z27116; CAAB1639.1; -				
DR	EMBL: Z28313; CAAB2167.1; -				
DR	PIR: S38166; S38166.				
DR	PIR: S39129; S39129.				
DR	SGD: S0001796; YKR088C.				
KW	Hypothetical protein; Transmembrane.				
FT	TRANSMEM 56 115				POTENTIAL.
FT	TRANSMEM 138 162				POTENTIAL.
FT	TRANSMEM 173 191				POTENTIAL.
FT	TRANSMEM 222 246				POTENTIAL.
FT	TRANSMEM 253 271				POTENTIAL.
FT	TRANSMEM 287 309				POTENTIAL.
SO	SEQUENCE 337 AA; 38311 MW; 7EA95DD4E5AF77FE CRC64;				
QY	Query Match	62.1%;	Score 41;	DB 1;	Length 337;
	Best Local Similarity	66.7%;	Pred. No. 4.1;		
DB	Matches 8; Conservative 1;	Mismatches 3;	Indels 0;	Gaps 0;	
	1 DNDISPFSGDQ 12				
	28 DNDIMPNSNGQ 39				

RESULT 2

```

VFIO_ECOLI STANDARD: PRT: 886 AA.
ID VFIO_ECOLI
AC P76594: 047320:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yfio.
GN yfio OR B2584.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA MEDLINE-97426617: PubMed-9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE OF 1-612 FROM N.A.
RC STRAIN-K12;
RA Nishimoto H., Saito N.;
RL Submitted (SEP-1995) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: SOME, IN THE N-TERMINAL, TO SUCCINYL-COA LIGASES.
CC -1- SIMILARITY: STRONG, TO M.VANNASCHIT MJ0590.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000344: AAC75637.1: -
DR EMBL: D64044: BAA10925.1: -
DR Ecogene: EG14224: yfio.
DR InterPro: IPR003781: COA_binding.
DR InterPro: IPR000182: GCN5acetyltransf.
DR Pfam: PF00583: Acetyltransf. 1.
DR Pfam: PF02629: COA_binding. 1.
DR Hypothetical protein: Complete proteome.
KW COMPLETE 506 507 ST-> VA (IN REF. 2).
FT SEQUENCE 886 AA: 97987 MW: 255944B9E2961251 CRC64;
SQ
Query Match 62.1%; Score 41; DB 1; Length 886;
Best Local Similarity 70.0%; Pred. No. 12;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Oy 3 DISPSGDCQ 12
Db 693 DISPSGDNE 702

```

```

RC STRAIN-H37RV;
RX MEDLINE-98295987: PubMed-9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: L-aspartate 4-semialdehyde + pyruvate =
CC dihydrodipicolinate + 2 H(2)O.
CC -1- PATHWAY: BIOSYNTHESIS OF DIAMINOPIMELATE AND LYSINE FROM ASPARTATE
CC SEMIALDEHYDE, FIRST STEP.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE DHDPS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL008967: CNA15549.1: -
DR EMBL: AE007110: AKA47142.1: -
DR HSP: P05640: IDHP.
DR TIGR: MT2823: -
DR Tuberculist: RV2753c: -
DR InterPro: IPR002220: DHDPS.
DR InterPro: IPR005263: DAPA.
DR Pfam: PF00701: DHDPS. 1.
DR PRINTS: PR00146: DHDPSNTFAS.
DR Prodom: PD001859: DHDPS. 1.
DR TIGRFAMs: TIGR00674: dapa. 1.
DR PROSITE: PS00655: DHDPS.1. 1.
DR PROSITE: PS00666: DHDPS.2. 1.
KW Lyase; Diaminopimelate biosynthesis; Lysine biosynthesis;
KW Complete proteome.
FT ACT_SITE 171 171
FT SEQUENCE 300 AA: 30858 MW: 710CB322661416367 CRC64;
SQ
Query Match 59.1%; Score 39; DB 1; Length 300;
Best Local Similarity 75.0%; Pred. No. 8.4;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Oy 4 ISPSGDCG 11
Db 20 VTFPSGDCG 27

```

```

RESULT 3
DAPA_MYCTU STANDARD: PRT: 300 AA.
ID DAPA_MYCTU
AC 033295;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dihydrodipicolinate synthase (EC 4.2.1.52) (DHDPS).
GN DAPA OR RV2753C OR MT2823 OR MTV002.18C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RESULT 4
QUEA_PSEAE STANDARD: PRT: 347 AA.
ID QUEA_PSEAE
AC 09HXH8;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

```

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE S-adenosylmethionine:trna ribosyltransferase-isomerase (EC 5.-.-.-)
DE (Queuosine biosynthesis protein quea)
GN QUESA OR PA3824.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Steiner C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garner R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RT Nature 406:959-964 (2000).
RL
CC -I- FUNCTION: Synthesizes OQ from preOQ in a single S-
CC adenosylmethionine-requiring step. The ribosyl moiety of Adomet is
CC transferred and isomerized to the epoxycyclopentane residue of OQ
CC (By similarity).
CC -I- PATHWAY: Queuosine biosynthesis.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -I- SIMILARITY: BELONGS TO THE QUESA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE004799; ANG07211.1; -.
DR InterPro: IPR003699; Queuosine_synth.
DR Pfam: PF02547; Queuosine_synth. 1.
DR TIGRFAMs: TIGR00113; quea. 1.
KW Queuosine biosynthesis; Transferase; Isomerase; Complete proteome.
SQ SEQUENCE 347 AA; 38160 MW; 48ECBF074C2BC589 CRC64;

Query Match 59.1%; Score 39; DB 1; Length 347;
Best Local Similarity 60.0%; Pred. No. 9.9;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 DNDISPFSGD 10
Db 263 DSELKPFSGD 272

RESULT 5
GUNI_USTMA STANDARD; PRT; 393 AA.
ID GUNI_USTMA
AC P54424;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase 1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase 1)
DE (Cellulase 1) (EG 1).
GN EGL1.
OS Ustilago maydis (Smut fungus).
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_TaxID=5270;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FB011;
RX MEDLINE=96145728; PubMed=8590631;
RX Schaefer F., Wanner G., Kahmann R.;
RT "Filament-specific expression of a cellulase gene in the dimorphic

```

```

RT fungus Ustilago maydis.";
RL Biol. Chem. Hoppe-Seyler 376:617-625 (1995).
CC -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- TISSUE SPECIFICITY: HYPMAL TIP.
CC -I- DEVELOPMENTAL STAGE: EXPRESSED IN FILAMENTOUS DIKARYON.
CC -I- PTM: MAY ALSO BE O-GLYCOSYLATED.
CC -I- SIMILARITY: BELONGS TO CELLULASE FAMILY K (FAMILY 45 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S81598; AAB36147.1; -.
DR HSPSP: P43316; 2ENG.
DR InterPro: IPR000334; GH_45.
DR Pfam: PF02015; Glyco_hydro_45; 1.
DR PROSITE: PS01140; GLYCOSYL_HYDROL_F45; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; signal; Glycoprotein.
FT SIGNAL 1 26
FT CHAIN 27 393
FT ACT_SITE 34 34
FT ACT_SITE 152 152
FT DOMAIN 270 385
FT CARBOHYD 343 343
FT SEQUENCE 393 AA; 39594 MW; 65C753C610CD6AD3 CRC64;

Query Match 59.1%; Score 39; DB 1; Length 393;
Best Local Similarity 77.8%; Pred. No. 11;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNDISPFSG 9
Db 239 DNTISPSG 247

RESULT 6
RRPB_CVMA5 STANDARD; PRT; 2733 AA.
ID RRPB_CVMA5
AC P16342;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE RNA-directed RNA polymerase (EC 2.7.7.48) (ORF1B).
GN F2.
OS Murine coronavirus MHV (strain A59).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=11142;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90245573; PubMed=2159623;
RX Bredenbeek P.J., Pachuk C.J., Noten A.F.H., Charite J., Luytjes W.,
RX Weiss S.R., Spaan W.J.M.;
RT "The primary structure and expression of the second open reading
RT frame of the polymerase gene of the coronavirus MHV-A59: a highly
RT conserved polymerase is expressed by an efficient ribosomal
RT frameshifting mechanism.";
RL Nucleic Acids Res. 18:1825-1832 (1990).
CC -I- FUNCTION: THE RNA DEPENDENT RNA POLYMERASE OF CORONAVIRUSES IS
CC A MULTIFUNCTIONAL PROTEIN: IT CONTAINS THE ACTIVITIES NECESSARY
CC FOR THE TRANSCRIPTION OF NEGATIVE STRANDED RNA, LEADER RNA,
CC SUBGENOMIC MRNAs AND PROGENY VIRION RNA.
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC [RNA] (N).
CC -I- MISCELLANEOUS: THIS PROTEIN MAY BE TRANSLATED AS A 1a-1b
CC POLYPROTEIN BY A RIBOSOMAL FRAMESHIFTING MECHANISM.

```

```

CC -1- SIMILARITY: TO THE AVIAN CORONAVIRUS IBV RNA POLYMERASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X51939; CAAB6202.1; ALT_SEQ.
DR PIR: S08652; S08652.
DR PIR: S15760; S15760.
KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT DOMAIN 622 824 POLYMERASE.
FT DOMAIN 944 1014 CYS/HIS-RICH.
FT DOMAIN 1218 1506 HELICASE.
FT NP_BIND 1220 1227 ATP (BY SIMILARITY).
SQ SEQUENCE 2733 AA; 309218 MW; F3ACC8BF20D20C41 CRC64;

Query Match 59.1%; Score 39; DB 1; Length 2733;
Best local similarity 60.0%; Pred. No. 1e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 DNDISFSGD 10
DB 2531 DNDVNPFSVD 2540

RESULT 7
IPNS_STRJU STANDARD; PRT; 329 AA.
ID IPNS_STRJU
AC P18286;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Isopenicillin N synthetase (EC 1.-.-.) (IPNS) (Isopenicillin N
DE synthetase).
GN PCBC.
OS Streptomyces jumonjiniensis.
OC Bacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1945;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=89112164; PubMed=3216857;
RA Shiffman D., Meyerech M., Jensen S.E., Cohen G., Aharonowitz Y.;
RT "Cloning and comparative sequence analysis of the gene coding for
RT isopenicillin N synthase in Streptomyces.";
RL Mol. Genet. 214:562-569(1988).
RN [2]
RX MUTAGENESIS OF HISTIDINE AND ASPARTIC ACID RESIDUES.
RX MEDLINE=96229923; PubMed=8639682;
RA Borovok I., Landman O., Kreisberg-Zakarin R., Aharonowitz Y.,
RA Cohen G.;
RT "Ferrous active site of isopenicillin N synthase: genetic and sequence
RT analysis of the endogenous ligands.";
RL Biochemistry 35:1981-1987(1996).
CC -1- FUNCTION: REMOVES, IN THE PRESENCE OF OXYGEN, 4 HYDROGEN ATOMS
CC FROM DELTA-L-(ALPHA-AMINODIPYL)-L-CYSTEINYL-D-VALINE (ACV) TO
CC FORM THE AZETIDINONE AND THIAZOLIDINE RINGS OF ISOPENICILLIN.
CC -1- PATHWAY: IRON AND ASCORBATE.
CC -1- CATALYZES: CENTRAL ROLE IN THE BIOSYNTHESIS OF PENICILLIN AND
CC CEPHALOSPORIN.
CC -1- SIMILARITY: BELONGS TO THE IRON/ASCORBATE-DEPENDENT FAMILY OF
CC OXIDOREDUCTASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M36687; AAA26772.1; -.
DR HSSP: P05326; IER0.
DR InterPro: IPR005123; 2OG-Feril_Oxy.
DR InterPro: IPR002283; IPN_synth.
DR InterPro: IPR002057; Isopen_N_synth.
DR Pfam: PF03171; 2OG-Feril_Oxy; 1.
DR PRINTS: PR00682; IPNSYNTHASE.
DR PROSITE: PS00185; IPNS_1; 1.
DR PROSITE: PS00186; IPNS_2; 1.
KW Antibiotic biosynthesis; Oxidoreductase; Iron; Vitamin C.
FT METAL 212 212 IRON.
FT METAL 214 214 IRON.
FT METAL 268 268 IRON.
SQ SEQUENCE 329 AA; 37305 MW; 9C8F1EB8FB8BDFC6 CRC64;

Query Match 56.1%; Score 37; DB 1; Length 329;
Best local similarity 87.5%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 DISFSGD 10
DB 14 DISPLSGD 21

RESULT 8
XYLA_ACTMI STANDARD; PRT; 393 AA.
ID XYLA_ACTMI
AC P12851;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Xylose isomerase (EC 5.3.1.5).
GN XYLA OR XI.
OS Actinoplanes missouriensis.
OC Bacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Micromonosporinae; Micromonosporaceae;
OX NCBI_TaxID=1866;
RN [1]
RX SEQUENCE FROM N.A.
RX STRAIN=DSM 43046;
RX MEDLINE=90016811; PubMed=2798103;
RA Amore R., Hollenberg C.P.;
RT "Xylose isomerase from Actinoplanes missouriensis: primary structure
RT of the gene and the protein.";
RL Nucleic Acids Res. 17:7515-7515(1989).
RN [2]
RX X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE=89184498; PubMed=3237716;
RA Rey F., Jenkins J., Janin J., Lasters I., Alard P., Claessens M.,
RA Mathysens G., Wodak S.J.;
RT "Structural analysis of the 2.8 A model of xylose isomerase from
RT Actinoplanes missouriensis.";
RL Proteins 4:165-172(1988).
RN [3]
RX X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RX MEDLINE=92304915; PubMed=1610791;
RA Jenkins J., Janin J., Rey F., Chidam M., van Tilbeurgh H.,
RA Lasters I., de Maeyer M., van Belle D., Wodak S.J., Lauwereys M.,
RA Staessens P., Mabret N.T., Snaawaert J., Mathysens G.,
RA Lambell A.-M.;
RT "Protein engineering of xylose (glucose) isomerase from Actinoplanes
RT missouriensis. 1. Crystallography and site-directed mutagenesis of
RT metal binding sites.";
RL Biochemistry 31:5449-5458(1992).
RN [4]
RX X-RAY CRYSTALLOGRAPHY (4.1 ANGSTROMS).
RX MEDLINE=99190841; PubMed=10089406;
RA Rant M., Shepard W., Fourme R., Kahn R.;
RT "Multivalue length anomalous solvent contrast (MASC): derivation of
RT envelope structure-factor amplitudes and comparison with model

```

```

RT values. ",
RL Acta Crystallogr. D 55:157-167(1999).
RN [5]
RP MUTAGENESIS.
RX MEDLINE=92304916; PubMed=1610792;
RA Lamber A.-M., Lauwereys M., Stanssens P., Mrabet N.T., Snaawaert J.,
RA van Tilbeurgh H., Malthysens G., Lasters I., de Maeyer M.,
RA Wodak S.J., Jenkins J., Chidam M., Janin J.;
RT "Protein engineering of xylose (glucose) isomerase from Actinoplanes
RT missouriensis. 3. Changing metal specificity and the pH profile by
RT site-directed mutagenesis. ";
RL Biochemistry 31:5467-5472(1992).
CC -1- CATALYTIC ACTIVITY: D-xylose = D-xyulose.
CC -1- COFACTOR: Magnesium: binds 2 ions per subunit.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE XYLOSE ISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16042; CAA34164.1; -;
DR EMBL; A10241; CAA00884.1; -;
DR PIR; S05998; ISMAXM.
DR PDB; 1XIN; 15-JUL-93.
DR PDB; 2XIN; 15-JUL-93.
DR PDB; 3XIN; 15-JUL-93.
DR PDB; 4XIM; 15-APR-93.
DR PDB; 5XIN; 15-JUL-93.
DR PDB; 6XIM; 15-JUL-93.
DR PDB; 7XIM; 15-JUL-93.
DR PDB; 8XIM; 15-JUL-93.
DR PDB; 9XIM; 15-JUL-93.
DR PDB; 1BHW; 18-NOV-98.
DR InterPro: IPR001998; xylose_isom.
DR Pfam; PF00259; xylose_isom; 1.
DR PROSITE; PRO00688; XYLOSISOMRASE.
DR PROSITE; PS00172; XYLOSE_ISOMERASE_1; 1.
DR PROSITE; PS00173; XYLOSE_ISOMERASE_2; 1.
KW Isomerase; Pentose shunt; Xylose metabolism; Magnesium; 3D-structure.
FT INT_MET 0
FT ACT_SITE 53
FT METAL 180
FT METAL 216
FT METAL 219
FT METAL 244
FT METAL 254
FT METAL 256
FT METAL 291
FT HELIX 6
FT STRAND 10
FT HELIX 13
FT TURN 14
FT TURN 18
FT TURN 23
FT TURN 24
FT TURN 25

```

FT	STRAND	26	26
FT	HELIX	35	45
FT	TURN	46	46
FT	STRAND	49	51
FT	STRAND	53	53
FT	HELIX	54	57
FT	TURN	60	61
FT	HELIX	64	81
FT	TURN	82	82
FT	STRAND	84	84
FT	STRAND	87	89
FT	HELIX	96	98
FT	TURN	99	100
FT	TURN	103	104
FT	HELIX	108	127
FT	TURN	128	129
FT	STRAND	132	136
FT	TURN	138	139
FT	STRAND	141	142
FT	HELIX	145	147
FT	HELIX	150	170
FT	TURN	171	172
FT	STRAND	176	180
FT	STRAND	189	190
FT	HELIX	195	202
FT	TURN	203	204
FT	HELIX	208	210
FT	STRAND	211	213
FT	STRAND	216	216
FT	HELIX	217	221
FT	TURN	222	224
FT	HELIX	227	237
FT	TURN	238	238
FT	STRAND	240	240
FT	STRAND	244	245
FT	STRAND	247	247
FT	STRAND	257	257
FT	TURN	259	260
FT	HELIX	264	275
FT	TURN	278	279
FT	STRAND	288	290
FT	TURN	296	297
FT	HELIX	300	326
FT	HELIX	328	336
FT	TURN	337	338
FT	HELIX	340	343
FT	TURN	349	350
FT	HELIX	353	358
FT	HELIX	360	362
FT	TURN	363	365
FT	HELIX	368	372
FT	TURN	373	373
FT	HELIX	378	389
FT	TURN	390	391
SO	SEQUENCE	393 AA;	43367 MM; 2BD72A4622C14A2A CRC64;
Query Match			
Best Local Similarity		56.1%;	Score 37; DB 1; Length 393;
Matches		6; Conservative	2; Mismatches 4; Indels 0; Gaps 0;
OY	1 DNDISPFSGDQ	12	
	1:1: 11 11		
Db	54 DDDLVFPGSDAQ	65	
RESULT 9			
ID	U120_HSVMG	STANDARD;	PRT; 402 AA.
AC	Q05101;		
DT	01-FEB-1995 (Rel. 31, Last sequence update)		
DT	01-FEB-1995 (Rel. 31, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Probable 44.7 kDa protein kinase (EC 2.7.1.-)		

GN US1206 OR US3.  
 OS Marek's disease herpesvirus (strain GA) (MDHV).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Marek's disease-like viruses.  
 OX NCBI\_TaxID=10388;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93118245; PubMed=1282282;  
 RA Sakauchi M., Urakawa T., Hirayama Y., Miki N., Yamamoto M.,  
 RA Hirai K.;  
 RT "Sequence determination and genetic content of an 8.9-kb restriction  
 RT fragment in the short unique region and the internal inverted repeat  
 RT of Marek's disease virus type 1 DNA."  
 RL Virus Genes 6:365-378(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9513166; PubMed=7831788;  
 RA Brnovskis P., Velicer L.F.;  
 RT "The Marek's disease virus (MDV) unique short region:  
 RT alphaherpesvirus-homologous, fowlpox virus-homologous, and  
 RT MDV-specific genes."  
 RL Virology 206:324-338(1995).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: M80595; AAB59895.1; -  
 CC DR EMBL: L22174; AAB64965.1; -  
 CC DR InterPro: IPR000719; Euk.pkinase.  
 CC DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 CC DR Pfam: PF00069; pkinase; 1.  
 CC DR ProDom: PD000001; Euk\_pkinase; 1.  
 CC DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
 CC DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 CC DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC KW Transferase: Serine/threonine-protein kinase; ATP-binding.  
 FT DOMAIN 102 386  
 FT NP\_BIND 110 118 ATP (BY SIMILARITY).  
 FT BINDING 127 127 ATP (BY SIMILARITY).  
 FT ACT\_SITE 218 218 BY SIMILARITY.  
 SQ SEQUENCE 402 AA: 44715 MW: 93E29494B3572C7B CRC64;  
 Query Match 56.1%; Score 37; DB 1; Length 402;  
 Best Local Similarity 66.7%; Pred. No. 27;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 OY 3 DISPSGDG 11  
 DB 74 DISPSGNDG 82  
 RESULT 10  
 MEK1\_YEAST  
 ID MEK1\_YEAST STANDARD: PRT: 497 AA.  
 AC P24719;  
 DT 01-MAR-1992 (rel. 21, Created)  
 DT 01-MAR-1992 (rel. 21, Last sequence update)  
 DT 16-OCT-2001 (rel. 40, Last annotation update)  
 DE Meiosis-specific serine/threonine-protein kinase MEK1 (EC 2.7.1.37).  
 GN MEK1 OR MBR4 OR YOR351C OR O6357.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92090720; PubMed=1752435;

RA Rockmill B., Roeder G.S.;  
 RT "A meiosis-specific protein kinase homolog required for chromosome  
 RT synapsis and recombination."  
 RL Genes Dev. 5:2392-2404(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92158649; PubMed=1741279;  
 RA Leem S.-H., Ogawa H.;  
 RT "The MRE1 gene encodes a novel protein kinase homolog required for  
 RT meiotic recombination in *Saccharomyces cerevisiae*."  
 RL Nucleic Acids Res. 20:449-457(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Delius H., Hebling U., Hofmann B.;  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SPROUCE OF 150-497 FROM N.A.  
 RC STRAIN=5288c / FV73;  
 RX MEDLINE=97103776; PubMed=8948102;  
 RA Purnelle B., Goffeau A.;  
 RT "Nucleotide sequence analysis of a 40 kb segment on the right arm of  
 RT yeast chromosome XY reveals 18 open reading frames including a new  
 RT pyruvate kinase and three homologues to chromosome I genes."  
 RL Yeast 12:1475-1481(1996).  
 CC -1- FUNCTION: PROBABLE PROTEIN KINASE REQUIRED FOR MEIOTIC  
 CC RECOMBINATION.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC -1- SIMILARITY: CONTAINS 1 FHA DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: X61208; CAA43522.1; -  
 CC DR EMBL: X63112; CAA44825.1; -  
 CC DR EMBL: X95720; CAA65038.1; -  
 CC DR EMBL: Z75259; CAA96880.1; -  
 CC DR PIR: A41637; A41637.  
 CC DR PIR: A41637; A41637.  
 CC DR PIR: S19065; S19065.  
 CC DR PIR: S20174; S20174.  
 CC DR HSSP: O63450; 1A06.  
 CC DR SGD: S0005878; MEK1.  
 CC DR InterPro: IPR000719; Euk.pkinase.  
 CC DR InterPro: IPR000253; FHA\_domain.  
 CC DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 CC DR Pfam: PF00069; pkinase; 1.  
 CC DR Pfam: PF00498; FHA; 1.  
 CC DR ProDom: PD000001; Euk.pkinase; 1.  
 CC DR SMART: SM00240; FHA; 1.  
 CC DR SMART: SM00220; S\_TKc; 1.  
 CC DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 CC DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC DR PROSITE: PS50006; FHA\_DOMAIN; 1.  
 CC KW Transferase: Serine/threonine-protein kinase; ATP-binding; Meiosis.  
 FT DOMAIN 47 102  
 FT NP\_BIND 162 162 FHA.  
 FT BINDING 199 199 ATP (BY SIMILARITY).  
 FT ACT\_SITE 290 290 ATP (BY SIMILARITY).  
 SQ SEQUENCE 497 AA: 56849 MW: 7E197338801EFD0 CRC64;  
 Query Match 56.1%; Score 37; DB 1; Length 497;  
 Best Local Similarity 87.5%; Pred. No. 35;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 4 ISPSGDG 11  
 I I I I I I I I

DB 382 ISPFYGDG 389

RESULT 11

ID	PNT1_DROME	STANDARD;	PRT;	623 AA.
AC	P51022; P19420;			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DE	ETS-1-like protein pointed P1 (D-ETS-2).			
GN	PNT OR ETS58AB OR ETS2 OR Cg17077.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;			
OC	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;			
OC	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxId=7227;			
RP	SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.			
RX	MEDLINE=94038653; PubMed=8223245;			
RT	Klaembt C.;			
RT	"The Drosophila gene pointed encodes two ETS-like proteins which are			
RL	involved in the development of the midline glial cells.";			
RL	Development 117:163-176(1993).			
RN	[2]			
RP	SEQUENCE OF 456-613 FROM N.A.			
RX	MEDLINE=92249640; PubMed=1577186;			
RA	Chen T., Bunting M., Karim F.D., Thummel C.S.;			
RT	"Isolation and characterization of five Drosophila genes that encode			
RT	an ets-related DNA binding domain.";			
RL	Dev. Biol. 151:176-191(1992).			
RN	[3]			
RP	SEQUENCE OF 445-603 FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL			
RP	STAGE.			
RX	MEDLINE=89196618; PubMed=2834248;			
RA	Pribyl L.J., Watson D.K., McWilliams M.J., Asclone R., Papas T.S.;			
RT	"The Drosophila ets-2 gene: molecular structure, chromosomal			
RT	localization, and developmental expression.";			
RL	Dev. Biol. 127:45-53(1988)			
CC	-1- FUNCTION: REQUIRED FOR GLIAL-NEURONAL CELL INTERACTIONS AT THE			
CC	VENERAL MIDLINE WHICH ARE NECESSARY FOR THE PROPER ELABORATION OF			
CC	COMMISSURES IN THE EMBRYONIC CNS.			
CC	-1- SUBCELLULAR LOCATION: Nuclear (Potential).			
CC	P51023); ARE PRODUCED BY ALTERNATIVE SPLICING.			
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN A COMPLEX DYNAMIC PATTERN IN			
CC	EARLY EMBRYOS, INCLUDING THE MIDLINE AND MIDLINE GLIAL CELLS.			
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT WITH LOWER			
CC	LEVELS DURING LARVAL DEVELOPMENT.			
CC	-1- SIMILARITY: BELONGS TO THE ETS FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; X69166; CAA48916.1; -			
DR	EMBL; M88472; AAC34200.1; -			
DR	EMBL; M20408; AAA28521.1; -			
DR	PIR; S28818; S28818.			
DR	HSSP; P14921; 2STT.			
DR	FLYbase; FBgn0003118; pnt.			
DR	InterPro; IPR000418; Ets.			
DR	InterPro; IPR002341; HSF_ETs.			
DR	Pfam; PF00178; Ets; 1			
DR	PRINTS; PR00454; ETSDOMAIN.			
DR	SMART; SM00413; ETS; 1.			
DR	PROSITE; PS00345; ETS_DOMAIN_1; 1.			
DR	PROSITE; PS00346; ETS_DOMAIN_2; 1.			
DR	PROSITE; PS50061; ETS_DOMAIN_3; 1.			

KW DNA-binding; Nuclear protein; Developmental protein;

KM Alternative splicing;

FT	DOMAIN	38	46	POLY-SER.
FT <td>DOMAIN</td> <td>47</td> <td>50</td> <td>POLY-ASN.</td>	DOMAIN	47	50	POLY-ASN.
FT <td>DOMAIN</td> <td>75</td> <td>88</td> <td>POLY-HIS.</td>	DOMAIN	75	88	POLY-HIS.
FT <td>DOMAIN</td> <td>103</td> <td>111</td> <td>POLY-GLN.</td>	DOMAIN	103	111	POLY-GLN.
FT <td>DOMAIN</td> <td>119</td> <td>124</td> <td>POLY-GLN.</td>	DOMAIN	119	124	POLY-GLN.
FT <td>DOMAIN</td> <td>128</td> <td>134</td> <td>POLY-THR.</td>	DOMAIN	128	134	POLY-THR.
FT <td>DOMAIN</td> <td>156</td> <td>159</td> <td>POLY-SER.</td>	DOMAIN	156	159	POLY-SER.
FT <td>DOMAIN</td> <td>160</td> <td>174</td> <td>POLY-SER.</td>	DOMAIN	160	174	POLY-SER.
FT <td>DOMAIN</td> <td>246</td> <td>252</td> <td>POLY-GLN.</td>	DOMAIN	246	252	POLY-GLN.
FT <td>DOMAIN</td> <td>264</td> <td>267</td> <td>POLY-ASN.</td>	DOMAIN	264	267	POLY-ASN.
FT <td>DOMAIN</td> <td>294</td> <td>298</td> <td>POLY-ASN.</td>	DOMAIN	294	298	POLY-ASN.
FT <td>DOMAIN</td> <td>310</td> <td>320</td> <td>POLY-ASN.</td>	DOMAIN	310	320	POLY-ASN.
FT <td>DOMAIN</td> <td>323</td> <td>326</td> <td>POLY-ALA.</td>	DOMAIN	323	326	POLY-ALA.
FT <td>DOMAIN</td> <td>386</td> <td>389</td> <td>POLY-GLY.</td>	DOMAIN	386	389	POLY-GLY.
FT <td>DOMAIN</td> <td>515</td> <td>595</td> <td>ETS-DOMAIN.</td>	DOMAIN	515	595	ETS-DOMAIN.
FT <td>DNA_BIND</td> <td>515</td> <td>595</td> <td></td>	DNA_BIND	515	595	
FT <td>CONFLICT</td> <td>467</td> <td>467</td> <td>Q -&gt; R (IN REF. 3).</td>	CONFLICT	467	467	Q -> R (IN REF. 3).
SO <td>SEQUENCE</td> <td>623 AA;</td> <td>66866 MW;</td> <td>F47083D80EB6480 CRC64;</td>	SEQUENCE	623 AA;	66866 MW;	F47083D80EB6480 CRC64;

QY 2 NDISPFGD 10

DB 205 NDMAPFGD 213

RESULT 12

ID	ITB5_PAPCY	STANDARD;	PRT;	655 AA.
AC	007441;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Integrin beta-5 (Fragment).			
GN	ITGB5.			
OS	Papio cynocephalus (Yellow baboon).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;			
OC	Cercopithecoidea; Papio.			
OX	NCBI_TaxId=9556;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94040831; PubMed=8224922;			
RA	Shoji M., Hayzer D.J., Kim T.M., Runge M.S., Hanson S.R.;			
RT	"Human and baboon integrin beta 5 subunit-encoding mRNAs have			
RT	alternative polyadenylation sites.";			
RL	Gene 133:307-308(1993).			
CC	-1- FUNCTION: INTEGRIN ALPHA-V/BETA-5 IS A RECEPTOR FOR FIBRONECTIN.			
CC	IT RECOGNIZES THE SEQUENCE R-G-D IT ITS LIGAND.			
CC	-1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-5			
CC	ASSOCIATES WITH ALPHA-V.			
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-1- PTM: THE CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN DISULFIDE			
CC	BONDS.			
CC	-1- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; L1231; AAI16866.1; -			
DR	HSSP; P05106; IUV2.			
DR	InterPro; IPR000561; EGF-like.			
DR	InterPro; IPR002369; Integrin_B.			
DR	InterPro; IPR001169; Integrin_beta_C.			

DR InterPro: IPR002035; VWF\_A.  
 DR Pfam: PF00362; Integrin\_B; 1.  
 DR ProDom: PD001811; Integrin\_B; 1.  
 DR SMART: SM00181; EGF; 1.  
 DR SMART: SM00187; INB; 1.  
 DR SMART: SM00327; VWA; 1.  
 DR PROSITE: PS00243; INTEGRIN\_BETA; 2.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN; 2.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN; 2.  
 DR Integrin: Cell adhesion; Receptor; Transmembrane; Glycoprotein;  
 KW Repeat: Extracellular matrix; Cytoskeleton.  
 FT NON\_TER 1  
 FT DOMAIN 1 575 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 576 598 POTENTIAL.  
 FT DOMAIN 599 655 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN <1 234 VWA-LIKE.  
 FT DOMAIN 457 621 4 CYSTEINE-RICH TANDEN REPEATS.  
 FT REPEAT 321 368 I.  
 FT REPEAT 369 410 II.  
 FT REPEAT 411 449 III.  
 FT REPEAT 450 486 IV.  
 FT CARBOHYD 203 203 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 316 316 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 408 408 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 442 442 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 510 510 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 561 561 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 SQ SEQUENCE 655 AA; 72466 MW; BAC3A159CBE1596 CRC64;

Query Match 56.1%; Score 37; DB 1; Length 655;  
 Best Local Similarity 87.5%; Pred. No. 47;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 DNDISF8  
 Db 40 DNDISF8 47

RESULT 13  
 ID Y012\_CAEEL STANDARD; PRT; 696 AA.  
 AC P34668;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Putative ATP-dependent RNA helicase ZK686.2 in chromosome III.  
 GN ZK686.2.  
 OS Caenorhabditis elegans.  
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 CC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Bristol N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Alnscough R., Anderson K., Baynes C., Beirs M.,  
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnson L., Jones M., Kersey J., Kirsten J., Laisner N.,  
 RA Lavelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,  
 RA Sims M., Smailson N., Smith A., Smith M., Sonnenhammer E., Staden K.,  
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Waterston R., Watson A., Weinstein L., Wilkinson-Sproat J.,  
 RA Wooldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 CC -1- FUNCTION: PROBABLE ATP-BINDING RNA HELICASE.  
 CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L17337; AAA28223.1; -  
 CC PIR: S44912; S44912.  
 CC WormPep: ZK686.2; CE00456.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR000629; DEAD\_Box.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR Pfam: PF00270; DEAD; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR SMART: SM00490; HELIC\_C; 1.  
 DR PROSITE: PS00039; DEAD\_ATP\_HELICASE; 1.  
 KW Hypothetical protein; Helicase; ATP-binding; RNA-binding.  
 FT NP\_BIND 266 273 ATP (BY SIMILARITY).  
 FT SITE 377 380 DEAD BOX.  
 SQ SEQUENCE 696 AA; 78376 MW; C6E6F97FA47FA2C CRC64;

Query Match 56.1%; Score 37; DB 1; Length 696;  
 Best Local Similarity 54.5%; Pred. No. 51;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 DNDISFSGDG 11  
 Db 118 DDEIFPANDG 128

RESULT 14  
 ID KRO8\_YEAST STANDARD; PRT; 734 AA.  
 AC P36004;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable serine/threonine-protein kinase YKL168C (EC 2.7.1.-).  
 GN YKL168C OR YKL632.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE=94378719; PubMed=8091858;  
 RA Vandenbol M., Bolle P.-A., Dion C., Portetelle D., Hliger F.;  
 RT "Sequencing and analysis of a 20.5 kb DNA segment located on the left  
 RT arm of yeast chromosome XI.";  
 RL Yeast 10:525-533(1994).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z26878; CA81519.1; -  
 CC EMBL: Z28168; CA82010.1; -  
 DR PIR: S37998; S37998.  
 DR PIR: S38413; S38413.  
 DR PIR: S44593; S44593.  
 DR SGD: S0001651; YKL168C.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR ProDom: PD000001; Euk\_pkinase; 1.



```

DR DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein; transferase; Serine/threonine-protein kinase;
KW ATP-binding.
FT DOMAIN 422 722 PROTEIN KINASE.
FT NP_BIND 428 436 ATP (BY SIMILARITY).
FT BINDING 465 465 ATP (BY SIMILARITY).
FT ACT_SITE 573 573 BY SIMILARITY.
SQ SEQUENCE 734 AA; 83656 MW; 896D26B0600A917D CRC64;

Query Match 56.1%; Score 37; DB 1; Length 734;
Best Local Similarity 41.7%; Pred. NO. 54;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 DNDISPFSGDQ 12
Db 184 DNDLTPRYGSKN 195

RESULT 15
ITB5_MOUSE STANDARD: PRT: 798 AA.
ID ITB5_MOUSE
AC 070309; 070308; 088347;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin beta-5 precursor.
GN ITGB5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A. (ISOFORMS BETA-5A AND BETA-5B).
RC TISSUE=Liver;
RX MEDLINE=96196405; PubMed=9531507;
RA Zhang H., Tan S.M., Lu J.;
RT "cDNA cloning reveals two mouse betas integrin transcripts distinct in
RT cytoplasmic domains as a result of alternative splicing.";
RL Biochem. J. 331:631-637(1998).
[2]
RN SEQUENCE FROM N.A. (ISOFORM BETA-5A).
RP TISSUE=Brain;
RC MEDLINE=99098874; PubMed=9880508;
RA Feng X., Teitelbaum S.L., Quirz M.E., Towler D.A., Ross F.P.;
RT "Cloning of the murine betas integrin subunit promoter. Identification
RT of a novel sequence mediating granulocyte-macrophage colony-
RT stimulating factor-dependent repression of betas integrin gene
transcription.";
RL J. Biol. Chem. 274:1366-1374(1999).
[3]
CC -1- FUNCTION: INTEGRIN ALPHA-V/BETA-5 IS A RECEPTOR FOR FIBRONECTIN.
CC IT RECOGNIZES THE SEQUENCE R-G-D IT ITS LIGAND.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-5
CC ASSOCIATES WITH ALPHA-V.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: BETA-5A (SHOWN HERE) AND BETA-
CC 5B: ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 WFA-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF043257; AAC40110.1; -
DR EMBL: AF043256; AAC40109.1; -
DR EMBL: AF022110; AAD08782.1; -
DR HSSP: P05106.1;Iv2.

```

DR	MGD; MGI:96614; Itgb5.
DR	InterPro; IPR000361; EGF-like.
DR	InterPro; IPR002369; Integrin_B.
DR	InterPro; IPR001169; Integrin_beta_C.
DR	InterPro; IPR003659; plexin-like.
DR	InterPro; IPR002035; VWF_A.
DR	pfam; PF00362; integrin_B_1.
DR	PRINTS; PR01186; INTEGRINB.
DR	PfDomM; PD001811; Integrin_B_1.
DR	SMART; sm00181; EGF_1.
DR	SMART; sm00187; INB_1.
DR	SMART; sm00423; PST_1.
DR	SMART; sm00327; VWI_1.
DR	PROSITE; PS00243; INTEGRIN.BETA.2.
DR	PROSITE; PS00022; EGF_1; UNKNOWN.2.
DR	PROSITE; PS01186; EGF_2; UNKNOWN.2.
KW	Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
KV	Repeat; Signal; Alternative splicing.
FT	SIGNAL
FT	CHAIN
FT	DOMAIN
FT	TRANSSEM
FT	DOMAIN
FT	DOMAIN
FT	REPEAT
FT	REPEAT
FT	REPEAT
FT	DISELID
FT	DISELID
FT	DISELID
FT	DISELID
FT	DISELID
FT	DISELID
FT	DISELID
FT	DISELID
FT	DISELID
FT	DISELID
FT	DISELID
FT	DISELID
FT	DISELID
FT	DISELID
FT	DISELID
FT	DISELID
FT	DISELID
FT	CARBOHYD
FT	CARBOHYD
FT	VARSPLIC
FT	CONFLICT
FT	CONFLICT
FT	SEQUENCE

Query Match  
Score 37;  
DB 1;  
length 798;

Best Local Similarity 87.5%; Pred. No. 59;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 DNDISPS 8  
| | | | |  
Db 184 DKDISPS 191

Search completed: March 18, 2003, 08:30:14  
Job time : 10.5 secs

GenCore version 5.1.4.P5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OK protein - protein search, using sw model

Run on: March 18, 2003, 08:28:17 ; Search time 26 Seconds  
(without alignments)  
95.099 Million cell updates/sec

Title: US-09-812-485A-42

Perfect score: 66

Sequence: 1 DNDISPSGDGQ 12

Scoring table:

BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: SPREMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	66	100.0	525	4 Q9NQ76	Q9nq76 homo sapien
2	63	95.5	500	6 Q9SKG5	Q9SKG5 macaca fasc
3	63	95.5	555	6 Q9GML3	Q9GML3 macaca fasc
4	63	95.5	555	6 Q9N076	Q9N076 macaca fasc
5	60	90.9	433	11 Q92411	Q92411 mus musculu
6	60	90.9	435	11 Q9ES02	Q9ES02 ratu muscu
7	45	68.2	913	16 Q9P003	Q9P003 ureaplasma
8	43	65.2	629	3 Q9Y819	Q9Y819 schizosacch
9	42	63.6	420	16 Q9YZ01	Q9YZ01 anabaena sp
10	40	60.6	536	12 Q8VA07	Q8VA07 chikungunya
11	40	60.6	726	16 Q8XR29	Q8XR29 ralsstonia s
12	40	60.6	2732	12 Q9J3F2	Q9J3F2 murine hepa
13	40	60.6	2733	12 Q9PYA2	Q9PYA2 murine hepa
14	39	59.1	131	13 Q90W20	Q90W20 hippoglossu
15	39	59.1	207	11 Q8V1K4	Q8V1K4 mus musculu
16	39	59.1	237	11 Q91ZW8	Q91ZW8 mus musculu

17	39	59.1	434	3 Q9P456	Q9P456 aspergillus
18	39	59.1	466	3 Q93785	Q93785 trichoderma
19	39	59.1	2733	12 Q9J3E8	Q9J3E8 murine hepa
20	39	59.1	2733	12 Q9J3E8	Q9J3E8 murine hepa
21	38.5	58.3	793	5 Q9VYX0	Q9VYX0 drosophila
22	38	57.6	104	10 Q9LGC8	Q9LGC8 oryza sativ
23	38	57.6	338	10 Q9FKN7	Q9FKN7 arabidopsis
24	38	57.6	348	16 Q8X2E2	Q8X2E2 escherichia
25	38	57.6	444	16 Q9KJ7	Q9KJ7 bacillus ha
26	38	57.6	459	11 Q9D852	Q9D852 mus musculu
27	38	57.6	480	10 Q9XE29	Q9XE29 oryza sativ
28	38	57.6	546	2 Q8RNY5	Q8RNY5 escherichia
29	38	57.6	673	10 Q9AGT5	Q9AGT5 oryza sativ
30	38	57.6	736	10 Q9AY58	Q9AY58 oryza sativ
31	38	57.6	886	16 Q8MX2	Q8MX2 salmonella
32	38	57.6	886	16 Q8Z4J4	Q8Z4J4 salmonella
33	38	57.6	886	16 Q8X9F6	Q8X9F6 escherichia
34	38	57.6	1967	12 Q10378	Q10378 rice yellow
35	37	56.1	233	10 Q48898	Q48898 agrostis st
36	37	56.1	354	2 Q9L6M9	Q9L6M9 pseudomonas
37	37	56.1	370	5 Q9VX05	Q9VX05 drosophila
38	37	56.1	389	5 Q9N5B5	Q9N5B5 caenorhabdi
39	37	56.1	402	12 Q9E6L8	Q9E6L8 turkey hepa
40	37	56.1	490	17 Q27541	Q27541 methanobact
41	37	56.1	500	11 Q9CWX5	Q9CWX5 mus musculu
42	37	56.1	500	11 Q91ZV7	Q91ZV7 mus musculu
43	37	56.1	501	10 Q93Z04	Q93Z04 arabidopsis
44	37	56.1	542	10 Q9SV40	Q9SV40 arabidopsis
45	37	56.1	640	16 Q9CKM9	Q9CKM9 pasteurella

## ALIGNMENTS

RESULT 1					
Q9NQ76		PRELIMINARY;	PRT;	525 AA.	
AC Q9NQ76;					
DT 01-OCT-2000 (TREMBLrel. 15, Created)					
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)					
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)					
DE Matrix extracellular phosphoglycoprotein precursor.					
GN MEPE.					
OS Homo sapiens (Human).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.					
OX NCBI_Taxid:9606;					
RN [1]					
RP SEQUENCE FROM N.A.					
RC TISSUE=BONE;					
RX MEDLINE=20399567; PubMed=10945470;					
RA Rowe P.S.N., De Zoysa P.A., Dong R., Wang H.R., White K.E.,					
RA Econs M.J., Oudet C.L.;					
RT "MEPE, a new gene expressed in bone marrow and tumors causing					
RT osteomalacia.";					
RL Genomics 67:54-68(2000).					
RN [2]					
RP SEQUENCE FROM N.A.					
RX MEDLINE=21309068; PubMed=11414762;					
RA Argiro L., Desbarats M., Glorieux F.H., Ecartot B.;					
RT "Mepe, the gene encoding a tumor-secreted protein in oncogenic					
RT hypophosphatemic osteomalacia, is expressed in bone.";					
RL Genomics 74:342-351(2001).					
DR EMBL; AJ276396; CAB97250.1;					
DR EMBL; AF325916; AAK70343.1;					
KW Signal.					
FT SIGNAL					
FT CHAIN	18	525	POTENTIAL.		
SQ SEQUENCE	525 AA;	58419 MM;	MATRIX EXTRACELLULAR PHOSPHOGLYCOPROTEIN.		
Query Match	100.0%;	Score 66;	DB 4;	Length 525;	
Best Local Similarity	100.0%;	Pred. No. 0.00072;			
Matches 12;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	

QY 1 DNDISPFSGDQ 12  
 DB 249 DNDISPFSGDQ 260

## RESULT 2

Q95KG5 PRELIMINARY; PRT; 500 AA.  
 ID 095KG5;  
 AC 095KG5;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DE Hypothetical 55.6 kDa protein.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Macaca.  
 RX NCBL\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TEMPORAL LOBE RIGHT;  
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,  
 RT Suzuki Y., Sugano S., Hashimoto K.;  
 RT "Isolation of full-length cDNA clones from macaque brain cDNA  
 libraries."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB060891; BAB46894.1; -  
 SO Hypothetical protein.  
 SQ SEQUENCE 500 AA; 55577 MW; 918D265AD8EDC7BC CRC64;

Query Match 95.5%; Score 63; DB 6; Length 500;  
 Best Local Similarity 91.7%; Pred. No. 0.0025;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DNDISPFSGDQ 12  
 DB 225 DNDISPFSGDQ 236

## RESULT 3

Q9GML3 PRELIMINARY; PRT; 555 AA.  
 ID 09GML3;  
 AC 09GML3;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE MEPE protein.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Macaca.  
 RX NCBL\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN PARITIAL LOBE;  
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,  
 RT Suzuki Y., Sugano S., Hashimoto K.;  
 RT "Isolation of full-length cDNA clones from macaque brain cDNA  
 libraries."  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB050259; BAB17010.1; -  
 SQ SEQUENCE 555 AA; 61950 MW; BD4D47E88038A9E2 CRC64;

Query Match 95.5%; Score 63; DB 6; Length 555;  
 Best Local Similarity 91.7%; Pred. No. 0.0028;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DNDISPFSGDQ 12  
 DB 280 DNDISPFSGDQ 291

RESULT 4  
 Q9N076 PRELIMINARY; PRT; 555 AA.  
 ID 09N076;  
 AC 09N076;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE Matrix extracellular phosphoglycoprotein.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Macaca.  
 RX NCBL\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,  
 RT Suzuki Y., Sugano S., Hashimoto K.;  
 RT "Isolation of full-length cDNA clones from macaque brain cDNA  
 libraries."  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB046056; BAB01638.1; -  
 SQ SEQUENCE 555 AA; 61979 MW; CCFELA98ADA19EB4 CRC64;

Query Match 95.5%; Score 63; DB 6; Length 555;  
 Best Local Similarity 91.7%; Pred. No. 0.0028;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DNDISPFSGDQ 12  
 DB 280 DNDISPFSGDQ 291

## RESULT 5

Q924I1 PRELIMINARY; PRT; 433 AA.  
 ID 0924I1;  
 AC 0924I1;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DE Matrix extracellular phosphoglycoprotein precursor.  
 GN MEPE.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RX NCBL\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RX MEDLINE=21309068; PubMed=11414762;  
 RA Argiro L., Desbarats M., Glorieux F.H., Ecarot B.;  
 RT "Mepe, the gene encoding a tumor-secreted protein in oncogenic  
 RT hypophosphatemic osteomalacia, is expressed in bone."  
 RL Genomics 74:342-351(2001).  
 DR EMBL; AF314964; AAK70342.1; -  
 DR MGD; MGI:2137384; Mepe.  
 KW Signal.  
 FT SIGNAL.  
 FT CHAIN.  
 SQ SEQUENCE 433 AA; 45984 MW; 7CD603CAACB41B0 CRC64;

Query Match 90.9%; Score 60; DB 11; Length 433;  
 Best Local Similarity 83.3%; Pred. No. 0.0077;  
 Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNDISPFSGDQ 12  
 DB 177 DNDISPFSGDQ 188

## RESULT 6

Q9ES02 PRELIMINARY; PRT; 435 AA.  
 ID 09ES02;  
 AC 09ES02;

DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE Osteoregulin.  
OC Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY;  
RX MEDLINE=20549633; PubMed=10967096;  
RA Petersen D.N., Tkalevic G.T., Mansolf A.L., Rivera-Gonzalez R.,  
RA Brown T.A.;  
RT "Identification of Osteoblast/Osteocyte Factor 45 (OF45), a Bone-  
RT specific cDNA Encoding an RGD-containing Protein that is Highly  
RT Expressed in Osteoblasts and Osteocytes";  
RL J. Biol. Chem. 275:36172-36180(2000).  
DR EMBL; AF260922; AAG33366.1; -.  
SQ SEQUENCE 435 AA; 46515 MW; D587F82968A26BCB CRC64;  
  
Query Match  
Best Local Similarity 90.9%; Score 60; DB 11; Length 435;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 NDISPSPSGDQ 12  
DB 171 DNDVPPFSGDQ 182

RESULT 7  
ID 09P003 PRELIMINARY; PRT; 913 AA.  
AC 09P003;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE MBA N-terminal paralog.  
GN U0487.  
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Ureaplasma.  
OX NCBI\_TaxID=134821;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SEROVAR 3;  
RX MEDLINE=20500219; PubMed=11048724;  
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,  
RA Cassell G.H.;  
RT "The complete sequence of the mucosal pathogen Ureaplasma  
RT urealyticum".  
RL Nature 407:757-762(2000).  
DR EMBL; AE002146; AAF30899.1; -.  
KW Complete proteome.  
SQ SEQUENCE 913 AA; 106010 MW; FE9AD007C44AE0E3 CRC64;

Query Match  
Best Local Similarity 68.2%; Score 45; DB 16; Length 913;  
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 NDISPSPSGDQ 12  
DB 891 NDISPSPSKDDQ 901

RESULT 8  
ID 09Y819 PRELIMINARY; PRT; 629 AA.  
AC 09Y819;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE Putative transmembrane protein.

GN SPBC1105.08.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972H-;  
RA McDougall R.C., Rajandream M.A., Barrell B.G., Ramsperger U.,  
RA Bothe G., Pohl T;  
RT Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AL096851; CAB50971.1; -.  
DR InterPro: IPR004240; EMP70.  
DR Pfam: PF02990; EMP70, 1.  
KW Transmembrane.  
SQ SEQUENCE 629 AA; 71130 MW; AC138984B757BAC2 CRC64;

Query Match  
Best Local Similarity 65.2%; Score 43; DB 3; Length 629;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 NDISPSPSGDQ 12  
DB 48 NTISPPIGDGR 58

RESULT 9  
ID 08Y201 PRELIMINARY; PRT; 420 AA.  
AC 08Y201;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Hypothetical protein A110406.  
GN A110406.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kurita T., Sasamoto S.,  
RA Matsumoto A., Iriguchi M., Ishikawa A., Kawashina K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
RT cyanobacterium Anabaena sp. strain PCC 7120".  
RL DNA Res. 8:205-213(2001).  
DR EMBL; AP003582; BAB72364.1; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 420 AA; 45452 MW; 036140FBE2513393 CRC64;

Query Match  
Best Local Similarity 70.0%; Score 42; DB 16; Length 420;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 NDISPSPSGDQ 11  
DB 60 NGINPFAGDG 69

RESULT 10  
ID 08VA07 PRELIMINARY; PRT; 536 AA.  
AC 08VA07;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Nonstructural protein 1 (Fragment).  
OS Chikungunya virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;

```

OC Alphavirus.
OX NCBI_TaxID=37124;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ROSS;
RA Logue C.H., Atkins G.J.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF438162; AL30897.1;
DR InterPro: IPR002588; Vmethyltransf.
DR Pfam: PF01660; Vmethyltransf; 1.
FT NON_TER
SQ SEQUENCE 536 AA; 59954 MW; 61B6FA0359D643F8 CRC64;

Query Match 60.6%; Score 40; DB 12; Length 536;
Best Local Similarity 58.3%; Pred. No. 55;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 DISPSGDDQ 12
   1::11111
Db 479 DLPYSGDAQ 488

RESULT 11
Q8XRZ9 PRELIMINARY; PRT; 726 AA.
ID Q8XRZ9
AC Q8XRZ9;
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Probable trimethylamine dehydrogenase oxidoreductase protein
   (EC 1.5.99.-).
GN RSP0682 OR RS01767.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM1000;
RX MEDLINE-21681879; PubMed-11823852;
RA Salanoubat M., Genin S., Artiguenave F., Guzy J., Mangenot S.,
RA Ariat M., Billault A., Brotier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siguier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL: AL646080; CAD17833.1;
DR InterPro: IPR000205; NAD_binding.
DR InterPro: IPR001155; Oxidored_FMN.
DR Pfam: PF00724; oxidored_FMN; 1.
KW Oxidoreductase; Plasmid; Complete proteome.
SQ SEQUENCE 726 AA; 79888 MW; 667F300EE328050C CRC64;

Query Match 60.6%; Score 40; DB 16; Length 726;
Best Local Similarity 58.3%; Pred. No. 77;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DNDISPSGDDQ 12
   1::11111
Db 256 DVLNPSFGDDQ 267

RESULT 12
Q9J3F2 PRELIMINARY; PRT; 2732 AA.
ID Q9J3F2
AC Q9J3F2;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE RNA-directed RNA polymerase.

```

```

OS murine hepatitis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=11138;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PENN 97-1;
RA Das Sarma J., Hingley S.T., Lai M.M.C., Weiss S.R., Lavi E.;
RT "Pathogenesis and sequence analysis of mouse hepatitis virus type 2;
RT an experimental model system of acute meningitis and hepatitis in
RT mice.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF208066; AAF69332.1;
DR RNA-directed RNA polymerase.
SQ SEQUENCE 2732 AA; 309134 MW; 46D0385DE231DC49 CRC64;

Query Match 60.6%; Score 40; DB 12; Length 2732;
Best Local Similarity 70.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 DNDISPSGDD 10
   1111111
Db 2531 DNDINPFVSD 2539

RESULT 13
Q9PYA2 PRELIMINARY; PRT; 2733 AA.
ID Q9PYA2
AC Q9PYA2;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE RNA-directed RNA polymerase.
GN ORF1B.
OS murine hepatitis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=11138;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MHV-2;
RA Das Sarma J., Hingley S.T., Lai M.M.C., Weiss S.R., Lavi E.;
RT "Pathogenesis and sequence analysis of mouse hepatitis virus type 2;
RT an experimental model system of acute meningitis and hepatitis in
RT mice.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ML-11;
RA Das Sarma J., Hingley S.T., Lai M.M.C., Weiss S.R., Lavi E.;
RT "Pathogenesis and sequence analysis of mouse hepatitis virus type 2;
RT an experimental model system of acute meningitis and hepatitis in
RT mice.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF201929; AAF19384.1;
DR EMBL: AF207902; AAF68920.1;
DR RNA-directed RNA polymerase.
SQ SEQUENCE 2733 AA; 309420 MW; D04F6457578EC1E7 CRC64;

Query Match 60.6%; Score 40; DB 12; Length 2733;
Best Local Similarity 70.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 DNDISPSGDD 10
   1111111
Db 2531 DNDINPFVSD 2540

RESULT 14
Q9QW20 PRELIMINARY; PRT; 131 AA.
ID Q9QW20
AC Q9QW20;
DT 01-DEC-2001 (TREMblrel. 19, Created)

```

```

DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE FSH-beta protein precursor.
GN FSH-BETA.
OS Hippoglossus hippoglossus (Atlantic halibut).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectidae; Pleuronectidae; Hippoglossus.
OX NCBI_Taxid=8267;
RN [1]
RP SEQUENCE FROM N.A.
RA Weltzien F.A., Kobayashi T., Andersson E., Norberg B., Andersen O.;
RT "Molecular characterization of FSH-beta, LH-beta, and alpha-subunit of
RT Atlantic halibut (Hippoglossus hippoglossus).";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ417768; CADI0501.1; -.
DR InterPro; IPR000359; Cys_knot.
DR Pfam; PF00007; Cys_knot; 1.
KW Signal.
FT SIGNAL.
SQ SEQUENCE 131 AA; 13965 MW; 290CEE379861A34F CRC64;

Query Match
Best Local Similarity 59.1%; Score 39; DB 13; Length 131;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 DNDISPFSGD 10
DB 105 DTDCSPFSGD 114

RESULT 15
O8VIK4
ID O8VIK4 PRELIMINARY; PRT; 207 AA.
AC O8VIK4;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE SIGNR3.
GN CD209D OR SIGNR3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RA Park C.G., Steinman R.M.;
RT "Alternatively Spliced Forms of Mouse DC-SIGN Homologs.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF440280; AAL33584.1; -.
DR MGD; MGI:2157947; CD209d.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; Lectin_C; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; UNKNOWN_1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
SQ SEQUENCE 207 AA; 23500 MW; 16FBAE847ECAD3FB CRC64;

Query Match
Best Local Similarity 59.1%; Score 39; DB 11; Length 207;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 DNDISPFSGD 11
DB 172 DEDCAFFSGDG 182

```

Search completed: March 18, 2003, 08:31:14  
 Job time : 29 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 08:27:02 ; Search time 32.5 Seconds  
(without alignments)  
49.200 Million cell updates/sec

Title: US-09-812-485a-43

Perfect score: 39

Sequence: 1 DXDXSFXGXGXQ 12

Scoring table:

BLOSUM62DX  
Gapop 10.0 , Gapept 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.GeneSeq-101002: \*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT: \*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT: \*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT: \*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT: \*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT: \*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT: \*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT: \*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT: \*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT: \*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT: \*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT: \*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT: \*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT: \*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT: \*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT: \*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT: \*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT: \*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT: \*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT: \*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT: \*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: \*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT: \*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	100.0	12	23	AAO20372
2	39	100.0	12	23	AAO20373
3	39	100.0	15	23	AAU93724
4	39	100.0	15	23	AAO20377
5	39	100.0	23	23	AAU93726
6	39	100.0	23	23	AAO20379
7	39	100.0	40	23	AAU93703
8	39	100.0	40	23	AAO20353
9	39	100.0	97	23	AAU93681
10	39	100.0	97	23	AAO20331

11	39	100.0	348	20	AAV39324	PSGen12 protein.
12	39	100.0	430	21	AAV53812	Amino acid sequenc
13	39	100.0	430	22	AAE62669	Truncated phosphat
14	39	100.0	509	22	AAE82922	Human osteoregulin
15	39	100.0	525	22	AAE82920	Human osteoregulin
16	39	100.0	525	22	AAE62689	Human osteoregulin
17	39	100.0	525	23	AAE80856	Phosphatoin polyP
18	39	100.0	525	23	AAE13227	Human oncogenetic os
19	39	100.0	540	22	AAE82923	Human osteoregulin
20	39	100.0	556	22	AAE82921	Human osteoregulin
21	36	92.3	328	22	ABE64344	Drosophila melanog
22	36	92.3	349	21	AAE14088	Arabidopsis thalia
23	36	92.3	349	21	AAE50445	Arabidopsis thalia
24	36	92.3	350	21	AAE14087	Arabidopsis thalia
25	36	92.3	350	21	AAE50444	Arabidopsis thalia
26	36	92.3	358	21	AAE07577	Arabidopsis thalia
27	36	92.3	504	22	AAU16220	Human novel secret
28	36	92.3	664	21	AAE29450	Arabidopsis thalia
29	36	92.3	665	21	AAE29449	Arabidopsis thalia
30	36	92.3	778	22	ABE60520	Drosophila melanog
31	36	92.3	778	22	ABE66160	Drosophila melanog
32	36	92.3	1004	22	AAE07164	Human caspase recr
33	36	92.3	1069	21	AAE29231	Human caspase recr
34	36	92.3	1104	21	AAE29230	Arabidopsis thalia
35	36	92.3	1133	21	AAE29229	Arabidopsis thalia
36	36	92.3	1391	21	AAE44664	Arabidopsis thalia
37	35	89.7	47	21	AAE57290	Candida albicans a
38	35	89.7	336	21	AAE15425	Arabidopsis thalia
39	35	89.7	367	21	AAE15425	Arabidopsis thalia
40	35	89.7	368	21	AAE15423	Arabidopsis thalia
41	35	89.7	384	22	AAU24559	Human olfactory re
42	35	89.7	384	23	AAE85179	G-coupled olfactor
43	35	89.7	432	22	AAE04682	Bacillus subtilis
44	35	89.7	450	22	AAU33718	Pseudomonas aerugi
45	35	89.7	531	21	AAE24871	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1	
AAO20372	
AAO20372 standard; peptide: 12 AA.	
AC	
AAO20372:	
DT	
31-MAY-2002 (first entry)	
XX	
DE	Peptide of a calcium binding motif #1.
XX	
KW	Bone growth; RGD motif; integrin binding motif; calcium binding motif;
KW	glycosaminoglycan binding motif; bone loss; renal phosphate excretion;
KW	alveolar; teeth; odontoblast; osteoclast; dental tissue; skeletal loss;
KW	weakness.
XX	
OS	Unidentified.
XX	
PN	WO200214360-A1.
PD	
21-FEB-2002.	
PF	
14-AUG-2001; 2001WO-US25542.	
XX	
16-AUG-2000; 2000US-0641034.	
PR	
19-MAR-2001; 2001US-0812485.	
XX	
PA	(BIGB-) BIG BEAR BIO INC.
XX	
PI	Kumagai Y, Blacher RW, Yoneda T;
XX	
DR	WPI; 2002-291971/33.
XX	
PT	New peptide compound useful for reducing bone loss, is capable of

PT		enhancing bone growth, and comprises an integrin binding motif,
XX		glycosaminoglycan binding motif or a calcium binding motif -
PS	Claim 5; Page 13; 50pp; English.	
CC	The invention relates to a peptide compound capable of enhancing bone	
CC	growth, and comprising 10-50 amino acids in a sequence, where the amino	
CC	acids are in D- or L- conformation and the sequence comprises a motif	
CC	selected from an integrin binding motif, a glycosaminoglycan binding	
CC	motif and a calcium binding motif. The peptide of the invention is useful	
CC	for reducing bone loss and for reducing renal phosphate excretion in an	
CC	individual. The peptide is useful for promoting regeneration of alveolar	
CC	bone and/or teeth, and increases the number and activity of odontoblasts	
CC	/osteoclasts that help form dental tissues. The peptide is also useful	
CC	for treating or preventing a condition associated with skeletal loss or	
CC	weakness. This sequence represents a peptide of the calcium binding motif	
CC	# 1 of the invention.	
XX		
SX	Sequence 12 AA:	
	Query Match	100.0%; Score 39; DB 23; Length 12;
	Best Local Similarity	50.0%; Pred. No. 1.9;
	Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;	
OY	1 DXDXSFXFGXXO 12	
	:   :   :   :   :	
Dd	1 DNDISFSGDGO 12	
RESULT 2		
AAO20373	ID AAO20373 standard; peptide: 12 AA.	
XX	AAO20373:	
AC	31-MAY-2002 (first entry)	
DT		
XX	Peptide of calcium binding motif #2.	
DE		
XX	Bone growth; RGD motif; integrin binding motif; calcium binding motif;	
KW	glycosaminoglycan binding motif; bone loss; renal phosphate excretion;	
KM	alveolar; teeth; odontoblast; osteoclast; dental tissue; skeletal loss;	
XX	weakness.	
OS	Unidentified.	
XX		
FH	Key Location/Qualifiers	
FT	Misc-difference 2	
FT	//label- Xaa	
FT	/note- "Xaa is unknown"	
FT	Misc-difference 4	
FT	//label- Xaa	
FT	/note- "Xaa is unknown"	
FT	Misc-difference 6	
FT	//label- Xaa	
FT	/note- "Xaa is unknown"	
FT	Misc-difference 8	
FT	//label- Xaa	
FT	/note- "Xaa is unknown"	
FT	Misc-difference 10	
FT	//label- Xaa	
FT	/note- "Xaa is unknown"	
FT	Misc-difference 11	
FT	//label- Xaa	
FT	/note- "Xaa is unknown"	
XX		
XX	WO200214360-A1.	
PN		
PD	21-FEB-2002.	
XX		
PF	14-AUG-2001; 2001MO-US25542.	
XX		
R	16-AUG-2000; 2000US-0641034.	

PR	19-MAR-2001; 2001US-0812485.
XX	
XX	(BIGB-) BIG BEAR BIO INC.
PA	
XX	Kumagai Y, Blacher RW, Yoneda T;
XX	
PI	WPI; 2002-291971/33.
DR	
XX	New peptide compound useful for reducing bone loss, is capable of
PT	enhancing bone growth, and comprises an integrin binding motif,
PT	glycosaminoglycan binding motif or a calcium binding motif -
XX	
PS	Claim 4; Page 13; 50pp; English.
XX	
CC	The invention relates to a peptide compound capable of enhancing bone
CC	growth, and comprising 10-50 amino acids in a sequence, where the amino
CC	acids are in D- or L- conformation and the sequence comprises a motif
CC	selected from an integrin binding motif, a glycosaminoglycan binding
CC	motif and a calcium binding motif. The peptide of the invention is useful
CC	for reducing bone loss and for reducing renal phosphate excretion in an
CC	individual. The peptide is useful for promoting regeneration of alveolar
CC	bone and/or teeth, and increases the number and activity of odontoblasts
CC	/osteoclasts that help form dental tissues. The peptide is also useful
CC	for treating or preventing a condition associated with skeletal loss or
CC	weakness. This sequence represents a peptide of the calcium binding motif
CC	#2 of the invention.
XX	
SO	Sequence 12 AA:

Query Match	100.0%;	Score 39;	DB 23;	Length 12;
Best Local Similarity	100.0%;	Pred. No. 1.9;		
Matches 12;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1 DDXSXFXGXQ 12			
Db	1 DDXSXFXGXQ 12			
RESULT 3				
ID	AAU93724	standard; peptide: 15 AA.		
XX	AAU93724;			
AC	AAU93724;			
XX				
DT	02-JUL-2002	(first entry)		
XX				
DE	Dental product bone growth enhancing peptide #44.			
XX				
KW	Dental product; toothpaste; mouthwash; dental floss; bone growth;			
KW	integrin binding motif; RGD; skeletal disease; dental disease; tooth;			
RW	alveolar bone growth; osteoblast; odontoblast; osteopathic.			
XX				
OS	Synthetic.			
XX				
EN	W0200213775-A1.			
XX				
PD	21-FEB-2002.			
XX				
PF	09-AUG-2001; 2001WO-US25101.			
XX				
PR	16-AUG-2000; 2000US-225879P.			
XX				
PA	(BIGB-) BIG BEAR BIO INC.			
XX				
PI	Yoneda T, Nomizu M, Kumagai Y;			
XX				
DR	WPI: 2002-329525/36.			
XX				
PT	Dental product useful for treating skeletal diseases e.g. dental			
XX	diseases comprises a base material and a compound comprising specific			
PT	amino acid in a sequence containing the integrin binding motif -			
XX				
VS	Claim 12; Page 21; 44pp; English.			

XX The present invention relates to dental products such as toothpastes,  
 CC mouthwash and dental floss comprising a base material and a compound  
 CC which promotes bone growth. Such compounds are peptide sequences  
 CC comprising 10-50 amino acids and containing an integrin binding  
 CC motif such as RGD in the D- or L- form, preferably the L-configuration.  
 CC The peptides of the invention are useful for treating or preventing  
 CC skeletal diseases such as dental disease. The peptides enhance tooth  
 CC and/or alveolar bone growth on areas where deterioration has occurred,  
 CC as well as the growth or recruiting of osteoblast or odontoblast cells  
 CC on the surface of the new skeletal growth. AAU93681-AAU93726 represent  
 CC bone growth enhancing peptides for use in dental products.  
 XX

SO Sequence 15 AA;

Query Match 100.0%; Score 39; DB 23; Length 15;  
 Best Local Similarity 50.0%; Pred. No. 2.4;  
 Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 DXDXSFXGXQ 12  
 I:|:|:|:|:|

DB 4 DNDISPFSGDQ 15

RESULT 4  
 AA020377 standard; peptide; 15 AA.

XX AA020377;  
 XX  
 XX  
 XX 31-MAY-2002 (first entry)  
 XX  
 DE C-terminal amidated synthesised peptide D-00004.

XX Bone growth; RGD motif; integrin binding motif; calcium binding motif;  
 KW glycosaminoglycan binding motif; bone loss; renal phosphate excretion;  
 KW alveolar; teeth; odontoblast; osteoclast; dental tissue; skeletal loss;  
 KW weakness; D00004.

XX Synthetic.

OS  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 15  
 FT /note="C-terminal amide"

XX WO200214360-A1.  
 XX  
 XX 21-FEB-2002.  
 XX  
 XX 14-AUG-2001; 2001WO-US25542.  
 XX  
 XX 16-AUG-2000; 2000US-0641034.  
 PR 19-MAR-2001; 2001US-0812485.  
 XX  
 XX (BIGB-) BIG BEAR BIO INC.  
 PA  
 XX Kumagai Y, Blacher RW, Yoneda T;  
 PI  
 XX WPI: 2002-291971/33.  
 DR  
 XX  
 PT New peptide compound useful for reducing bone loss, is capable of  
 PT enhancing bone growth, and comprises an integrin binding motif,  
 PT glycosaminoglycan binding motif or a calcium binding motif -  
 XX  
 XX Example 1; Page 23; 50pp; English.

XX The invention relates to a peptide compound capable of enhancing bone  
 CC growth, and comprising 10-50 amino acids in a sequence, where the amino  
 CC acids are in D- or L- conformation and the sequence comprises a motif  
 CC selected from an integrin binding motif, a glycosaminoglycan binding  
 CC motif and a calcium binding motif. The peptide of the invention is useful  
 CC for reducing bone loss and for reducing renal phosphate excretion in an  
 CC individual. The peptide is useful for promoting regeneration of alveolar

CC bone and/or teeth, and increases the number and activity of odontoblasts  
 CC /osteoclasts that help form dental tissues. The peptide is also useful  
 CC for treating or preventing a condition associated with skeletal loss or  
 CC weakness. This sequence represents a C-terminal amidated synthesised  
 CC peptide D-00004 of the invention.  
 XX

SO Sequence 15 AA;

Query Match 100.0%; Score 39; DB 23; Length 15;  
 Best Local Similarity 50.0%; Pred. No. 2.4;  
 Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 DXDXSFXGXQ 12  
 I:|:|:|:|:|

DB 4 DNDISPFSGDQ 15

RESULT 5  
 AAU93726 standard; peptide; 23 AA.

XX AAU93726;  
 XX  
 XX  
 XX 02-JUL-2002 (first entry)  
 XX  
 DE Dental product bone growth enhancing peptide #46.

XX Dental product; toothpaste; mouthwash; dental floss; bone growth;  
 KW integrin binding motif; RGD; skeletal disease; dental disease; tooth;  
 KW alveolar bone growth; osteoblast; odontoblast; osteopathic.

OS Synthetic.

XX  
 XX  
 XX WO200213775-A1.  
 XX  
 XX 21-FEB-2002.  
 XX  
 XX 09-AUG-2001; 2001WO-US25101.  
 PF  
 XX 16-AUG-2000; 2000US-225879P.  
 PR  
 XX (BIGB-) BIG BEAR BIO INC.  
 PA  
 XX Yoneda T, Nomizu M, Kumagai Y;  
 PI  
 XX WPI: 2002-329525/36.  
 DR  
 XX  
 XX Claim 7; Page 21; 44pp; English.

XX The present invention relates to dental products such as toothpastes,  
 CC mouthwash and dental floss comprising a base material and a compound  
 CC which promotes bone growth. Such compounds are peptide sequences  
 CC comprising 10-50 amino acids and containing an integrin binding  
 CC motif such as RGD in the D- or L- form, preferably the L-configuration.  
 CC The peptides of the invention are useful for treating or preventing  
 CC skeletal diseases such as dental disease. The peptides enhance tooth  
 CC and/or alveolar bone growth on areas where deterioration has occurred,  
 CC as well as the growth or recruiting of osteoblast or odontoblast cells  
 CC on the surface of the new skeletal growth. AAU93681-AAU93726 represent  
 CC bone growth enhancing peptides for use in dental products.  
 XX

SO Sequence 23 AA;

Query Match 100.0%; Score 39; DB 23; Length 23;  
 Best Local Similarity 50.0%; Pred. No. 4;  
 Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 DXDXSFXGXQ 12  
 I:|:|:|:|:|

DB 8 DNDISPFSGDQ 19

RESULT 6  
AAO20379  
ID AAO20379 standard; peptide: 23 AA.  
XX  
AC AAO20379;  
XX  
DT 31-MAY-2002 (first entry)  
XX  
DE C-terminal amidated synthesised peptide D-00006.  
XX  
KW Bone growth; RGD motif; integrin binding motif; calcium binding motif;  
KW glycosaminoglycan binding motif; bone loss; renal phosphate excretion;  
KW alveolar; teeth; odontoblast; osteoclast; dental tissue; skeletal loss;  
KW weakness; D00006.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 23 /note="C-terminal amide"  
FT  
XX  
XX WO200214360-A1.  
XX  
PD 21-FEB-2002.  
XX  
PF 14-AUG-2001; 2001WO-US25542.  
XX  
PR 16-AUG-2000; 2000US-0641034.  
PR 19-MAR-2001; 2001US-0812485.  
XX  
XX (BIGB-) BIG BEAR BIO INC.  
XX  
PI Kumagai Y, Blacher RM, Yoneda T;  
XX  
DR WPI; 2002-291971/33.  
XX  
PT New peptide compound useful for reducing bone loss, is capable of  
PT enhancing bone growth, and comprises an integrin binding motif,  
PT glycosaminoglycan binding motif or a calcium binding motif -  
XX  
XX Example 1; Page 15; 50pp; English.  
XX  
PS The invention relates to a peptide compound capable of enhancing bone  
XX growth, and comprising 10-50 amino acids in a sequence, where the amino  
XX acids are in D- or L- conformation and the sequence comprises a motif  
XX selected from an integrin binding motif, a glycosaminoglycan binding  
XX motif and a calcium binding motif. The peptide of the invention is useful  
XX for reducing bone loss and for promoting renal phosphate excretion in an  
XX individual. The peptide is useful for promoting regeneration of alveolar  
XX bone and/or teeth, and increases the number and activity of odontoblasts  
XX /osteoclasts that help form dental tissues. The peptide is also useful  
XX for treating or preventing a condition associated with skeletal loss or  
XX weakness. This sequence represents a C-terminal amidated synthesised  
XX peptide D-00006 of the invention.  
XX  
SQ Sequence 23 AA;

Query Match 100.0%; Score 39; DB 23; Length 23;  
Best Local Similarity 50.0%; Pred. No. 4;  
Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 DXDXSFXGXXQ 12  
|:|:|:|:|:|:|  
Db 8 DNDISPFSGDQ 19

RESULT 7  
AAU93703  
ID AAU93703 standard; peptide: 40 AA.  
XX

AC AAU93703;  
XX  
DT 02-JUL-2002 (first entry)  
XX  
DE Dental product bone growth enhancing peptide #23.  
XX  
XX Dental product; toothpaste; mouthwash; dental floss; bone growth;  
KW integrin binding motif; RGD; skeletal disease; dental disease; tooth;  
KW alveolar bone growth; osteoblast; odontoblast; osteopathic.  
XX  
OS Synthetic.  
XX  
PN WO200213775-A1.  
XX  
PD 21-FEB-2002.  
XX  
PF 09-AUG-2001; 2001WO-US25101.  
XX  
PR 16-AUG-2000; 2000US-225879P.  
XX  
XX (BIGB-) BIG BEAR BIO INC.  
XX  
PI Yoneda T, Nomizu M, Kumagai Y;  
XX  
DR WPI; 2002-329525/36.  
XX  
PT Dental product useful for treating skeletal diseases e.g. dental  
PT diseases comprises a base material and a compound comprising specific  
PT amino acid in a sequence containing the integrin binding motif -  
XX  
XX Claim 6; Page 20; 44pp; English.  
XX  
PS The present invention relates to dental products such as toothpastes,  
XX mouthwash and dental floss comprising a base material and a compound  
XX which promotes bone growth. Such compounds are peptide sequences  
XX comprising 10-50 amino acids and containing an integrin binding  
XX motif such as RGD in the D- or L- form, preferably the L-configuration.  
XX The peptides of the invention are useful for treating or preventing  
XX skeletal diseases such as dental disease. The peptides enhance tooth  
XX and/or alveolar bone growth on areas where deterioration has occurred,  
XX as well as the growth or recruiting of osteoblast or odontoblast cells  
XX on the surface of the new skeletal growth. AAU93681-AAU93726 represent  
XX bone growth enhancing peptides for use in dental products.  
XX  
SQ Sequence 40 AA;

Query Match 100.0%; Score 39; DB 23; Length 40;  
Best Local Similarity 50.0%; Pred. No. 7.7;  
Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 DXDXSFXGXXQ 12  
|:|:|:|:|:|:|  
Db 3 DNDISPFSGDQ 14

RESULT 8  
AAO20353  
ID AAO20353 standard; protein: 40 AA.  
XX  
AC AAO20353;  
XX  
DT 31-MAY-2002 (first entry)  
XX  
DE Protein of matrix extracellular phosphoglycoprotein containing RGD #23.  
XX  
KW Bone growth; RGD motif; integrin binding motif; calcium binding motif;  
KW glycosaminoglycan binding motif; bone loss; renal phosphate excretion;  
KW alveolar; teeth; odontoblast; osteoclast; dental tissue; skeletal loss;  
KW weakness; matrix extracellular phosphoglycoprotein.  
XX  
OS Unidentified.  
XX  
PN WO200214360-A1.

XX 21-FEB-2002.  
PD  
XX  
PF 14-AUG-2001; 2001WO-US25542.  
XX  
PR 16-AUG-2000; 2000US-0641034.  
PR 19-MAR-2001; 2001US-0812485.  
XX  
PA (BIGB-) BIG BEAR BIO INC.  
XX  
PI Kumagai Y, Blacher RW, Yoneda T;  
XX WPI: 2002-291971/33.  
DR  
XX  
XX New peptide compound useful for reducing bone loss, is capable of  
PT enhancing bone growth, and comprises an integrin binding motif,  
PT glycosaminoglycan binding motif or a calcium binding motif -  
XX  
PS  
XX Disclosure: Page 12; 50pp; English.  
XX  
XX The invention relates to a peptide compound capable of enhancing bone  
CC growth, and comprising 10-50 amino acids in a sequence, where the amino  
CC acids are in D- or L- conformation and the sequence comprises a motif  
CC selected from an integrin binding motif, a glycosaminoglycan binding  
CC motif and a calcium binding motif. The peptide of the invention is useful  
CC for reducing bone loss and for reducing renal phosphate excretion in an  
CC individual. The peptide is useful for promoting regeneration of alveolar  
CC bone and/or teeth, and increases the number and activity of odontoblasts  
CC /osteoclasts that help form dental tissues. The peptide is also useful  
CC for treating or preventing a condition associated with skeletal loss or  
CC weakness. This sequence represents a protein of a matrix extracellular  
CC phosphoglycoprotein containing an RGD motif of the invention.  
XX  
XX Sequence 40 AA;

Query Match	100.0%;	Score 39;	DB 23;	Length 40;
Best local similarity	50.0%;	Pred. No. 7.7;		
Matches 6;	Conservative 6;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1 DDXSXFXGXGXQ 12			
	: : : : :			
Db	3 DNDISFSGDQ 14			
RESULT 9				
ID	AAU93681			
	AAU93681 standard; protein; 97 AA.			
XX	AAU93681;			
AC				
XX				
DT	02-JUL-2002 (first entry)			
XX				
DE	Dental product bone growth enhancing peptide #1.			
XX				
KW	Dental product; toothpaste; mouthwash; dental floss; bone growth;			
KM	integrin binding motif; RGD; skeletal disease; dental disease; tooth;			
KM	alveolar bone growth; osteoblast; odontoblast; osteopathic.			
XX				
OS	Synthetic.			
XX				
PN	WO200213775-A1.			
XX				
PD	21-FEB-2002.			
XX				
PF	09-AUG-2001; 2001WO-US25101.			
XX				
XX	16-AUG-2000; 2000US-225879P.			
XX				
PA	(BIGB-) BIG BEAR BIO INC.			
XX				
PI	Yoneda T, Nomizu M, Kumagai Y;			
XX				
WP	2002-329525/36.			
XX				

XX	Dental product useful for treating skeletal diseases e.g. dental
PT	diseases comprises a base material and a compound comprising specific
PT	amino acid in a sequence containing the integrin binding motif -
XX	
PS	Disclosure: Page 11; 44pp; English.
XX	
CC	The present invention relates to dental products such as toothpastes,
CC	mouthwash and dental floss comprising a base material and a compound
CC	which promotes bone growth. Such compounds are peptide sequences
CC	comprising 10-50 amino acids and containing an integrin binding
CC	motif such as RGD in the D- or L- form, preferably the L-configuration.
CC	The peptides of the invention are useful for treating or preventing
CC	skeletal diseases such as dental disease. The peptides enhance tooth
CC	and/or alveolar bone growth on areas where deterioration has occurred,
CC	as well as the growth or recruiting of osteoblast or odontoblast cells
CC	on the surface of the new skeletal growth. AA093681-AA093726 represent
CC	bone growth enhancing peptides for use in dental products.
XX	
SQ	Sequence 97 AA;
Query Match	100.0%; Score 39; DB 23; Length 97;
Best Local Similarity	50.0%; Pred. No. 22;
Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;	
QY	I DXDMSXFXGXGXXO 12  :: :: :: ::
Db	50 DNDISPFSGDGQ 6L
RESULT 10	
ID	AAO20331
XX	AAO20331 standard; protein; 97 AA.

RESULT 10	
AAO20331	
ID	AAO20331 standard; protein; 97 AA.
XX	
AC	
XX	AAO20331;
XX	
DT	31-MAY-2002 (first entry)
XX	
DE	Protein of matrix extracellular phosphoglycoprotein containing RGD #1.
XX	
XX	Bone growth; RGD motif; integrin binding motif; calcium binding motif;
RW	glycosaminoglycan binding motif; bone loss; renal phosphate excretion;
RW	alveolar; teeth; odontoblast; osteoclast; dental tissue; skeletal loss;
KW	weakness; matrix extracellular phosphoglycoprotein.
XX	
OS	Unidentified.
XX	
FN	MO200214360-A1.
XX	
PD	21-FEB-2002.
XX	
FE	14-AUG-2001; 2001WO-US25542.
XX	
PR	16-AUG-2000; 2000US-0641034.
PR	19-MAR-2001; 2001US-0812485.
XX	
PA	(BIGB-) BIG BEAR BIO INC.
XX	
P1	Kumagai Y, Blacher RW, Yoneda T;
XX	
DR	WPI; 2002-291971/33.
XX	
PT	New peptide compound useful for reducing bone loss, is capable of
PT	enhancing bone growth, and comprises an integrin binding motif,
PT	glycosaminoglycan binding motif or a calcium binding motif -
XX	
PS	
XX	Disclosure; Page 11; 50pp; English.
CC	
CC	The invention relates to a peptide compound capable of enhancing bone
CC	growth, and comprising 10-50 amino acids in a sequence, where the amino
CC	acids are in D- or L- conformation and the sequence comprises a motif
CC	selected from an integrin binding motif, a glycosaminoglycan binding
CC	motif and a calcium binding motif. The peptide of the invention is useful
CC	

CC for reducing bone loss and for reducing renal phosphate excretion in an  
CC individual. The peptide is useful for promoting regeneration of alveolar  
CC bone and/or teeth, and increases the number and activity of odontoblasts  
CC /osteoclasts that help form dental tissues. The peptide is also useful  
CC for treating or preventing a condition associated with skeletal loss or  
CC weakness. This sequence represents a protein of a matrix extracellular  
CC phosphoglycoprotein containing an RGD motif of the invention.

XX Sequence 97 AA;

Query Match 100.0%; Score 39; DB 23; Length 97;  
Best Local Similarity 50.0%; Pred. No. 22;  
Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDXSXFXGXQ 12  
|:|:|:|:|:|  
Db 50 DNDISFSGDQ 61

RESULT 11

AA539324  
ID AAY39324 standard; Protein; 348 AA.

XX AAY39324;

XX 01-DEC-1999 (first entry)

XX PSGen12 protein.

XX Progression suppressed gene; PSCen: progression elevated gene; PSCen:  
KM tumour; reciprocal subtraction differential RNA display; RSDd;  
KM differential expression; gene cloning; cancer.

OS Rattus sp.

XX MO9943844-A1.

XX 02-SEP-1999.

XX 26-FEB-1999; 99WO-US04323.

XX 27-FEB-1998; 98US-0032684.

XX 03-NOV-1998; 98US-0185115.

XX 23-NOV-1998; 98US-0197889.

XX (UYCO ) UNIV COLUMBIA NEW YORK.

XX Fisher PB;

XX WPI. 1999-550872/46.

XX N-PSDB: AA621516.

XX Identifying nucleic acids differentially expressed between two samples,  
XX particularly sequences involved in tumour progression -

XX Claim 40; Fig 35A; 110pp: English.

XX This is the amino acid sequence of the PSCen12 protein (progression  
CC suppressed gene 12). PSCen12 has suppressed expression in progressed  
CC tumour cells. The PSCen12 nucleic acid sequence was identified using new  
CC methods for identifying nucleic acids differentially expressed between  
CC two samples. The method involves performing reciprocal subtraction  
CC differential RNA display (RSDd) between the two samples to generate two  
CC subtraction samples. The subtraction samples are amplified and compared  
CC to identify those nucleic acids that are differentially expressed. The  
CC method is used to identify and clone differentially expressed genes,  
CC particularly those with increased or reduced expression during tumour  
CC cell progression, e.g. progression suppressed genes (PSCen) and  
CC progression elevated genes (PEGen). The method reduces the complexity of  
CC the band pattern produced in conventional differential RNA display (where  
CC bands may be obscured, resulting in false positive signals) since most  
CC bands common to both samples are eliminated, allowing identification and  
CC cloning of genes displaying anticipated differential expression. RSDd

CC requires only a single anchored primer for amplification and reamplified  
CC cDNA can be analysed by reverse Northern blotting.

XX Sequence 348 AA;

Query Match 100.0%; Score 39; DB 20; Length 348;  
Best Local Similarity 50.0%; Pred. No. 99;  
Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDXSXFXGXQ 12  
|:|:|:|:|:|  
Db 212 DFDKSFACRCQ 223

RESULT 12

AA53812  
ID AAY53812 standard; Protein; 430 AA.

XX AAY53812;

XX 22-FEB-2000 (first entry)

XX Amino acid sequence of a human phosphatonin polypeptide.

XX Human; phosphatonin; Metastatic-tumour Excreted Phosphaturic-Element;  
KM MEPE; Na+-dependent phosphate cotransport; vitamin D metabolism;  
KM bone mineralisation; phosphate metabolism related disease;

XX hyperphosphatemia; renal osteodystrophy; renal dialysis;  
KM secondary hyperparathyroidism; osteitis fibrosa cystica; hypercalcaemia;  
KM X-linked hypophosphatemic rickets; hereditary hypophosphatemic rickets;  
KM hypomineralised bone lesion; stunted growth; cystic fibrosis;

XX oncogenic hypophosphatemic osteomalacia; renal phosphate leakage;  
KM renal osteodystrophy; osteoporosis; vitamin D resistant rickets;  
KM end organ resistance; renal Fanconi syndrome; autosomal rickets;  
KM Paget's disease; kidney failure; renal tubular acidosis; spure.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Modified-site 8..10 /note= "protein kinase C phosphorylation site"

XX Modified-site 8..11 /note= "protein kinase II phosphorylation site"

XX Modified-site 16..21 /note= "myristoylation site"

XX Modified-site 40..47 /note= "Tyrosine kinase phosphorylation site"

XX Modified-site 77..79 /note= "protein kinase C phosphorylation site"

XX Modified-site 118..120 /note= "protein kinase C phosphorylation site"

XX Modified-site 119..124 /note= "myristoylation site"

XX Modified-site 139..142 /note= "Casein kinase II phosphorylation site"

XX Modified-site 143..148 /note= "myristoylation site"

XX Modified-site 152..154 /note= "cell attachment tripeptide"

XX Modified-site 161..165 /note= "glycosaminoglycan attachment site"

XX Modified-site 177..180 /note= "Casein kinase II phosphorylation site"

XX Modified-site 194..197 /note= "Casein kinase II phosphorylation site"

XX Modified-site 199..202 /note= "Casein kinase II phosphorylation site"

XX Modified-site 203..205 /note= "Casein kinase II phosphorylation site"

XX Modified-site 224..227 /note= "Casein kinase II phosphorylation site"

XX Modified-site 228..231 /note= "Casein kinase II phosphorylation site"

FT	Modified-site	228..230	/note= "protein kinase C phosphorylation site"
FT	Modified-site	238..241	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	266..271	/note= "myristoylation site"
FT	Modified-site	291..296	/note= "myristoylation site"
FT	Modified-site	311..313	/note= "protein kinase C phosphorylation site"
FT	Modified-site	312..314	/note= "protein kinase C phosphorylation site"
FT	Modified-site	315..320	/note= "myristoylation site"
FT	Modified-site	319..321	/note= "protein kinase C phosphorylation site"
FT	Modified-site	325..328	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	370..373	/note= "amidation site"
FT	Modified-site	382..386	/note= "Asu-glycosylation site"
FT	Modified-site	383..387	/note= "Asu-glycosylation site"
FT	Modified-site	384..386	/note= "protein kinase C phosphorylation site"
FT	Modified-site	389..394	/note= "myristoylation site"
FT	Modified-site	403..405	/note= "protein kinase C phosphorylation site"
FT	Modified-site	405..408	/note= "cAMP and cGMP dependent protlen kinase phosphorylation site"
FT	Modified-site	408..410	/note= "protein kinase C phosphorylation site"
FT	Modified-site	409..411	/note= "protein kinase C phosphorylation site"
FT	Modified-site	423..426	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	425..428	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	427..430	/note= "Casein kinase II phosphorylation site"
XX	WO9960017-A2.		
PD	25-NOV-1999.		
XX	18-MAY-1999;	99WO-EP03403.	
PR	18-MAY-1998;	98GB-0010681.	
PR	04-SEP-1998;	98GB-0019387.	
XX	(UNLO ) UNIV COLLEGE LONDON.		
XX	Rowe P;		
PI	WPI: 2000-053262/04.		
DR	N-PSDB: AA236447.		
XX	New polypeptides involved in the regulation of phosphate metabolism useful for diagnosing and treating disorders related to phosphate metabolism		
PS	Claim 6; Fig 8; 136pp: English.		
XX	The present sequence represents a phosphatonin polypeptide (also called Metastatic-tumour Excreted Phosphaturic-Element (MEPE)). The level of phosphatonin in a subject modulates Na+-dependent phosphate cotransport, vitamin D metabolism and/or bone mineralisation. The phosphatonin polypeptides, polynucleotides, vectors and antibodies are used to treat phosphate metabolism related disease. They are used for treatment of hyperphosphatemia, or renal osteodystrophy, hyperphosphatemia in renal		

[illegible]

CC disorder of phosphate or vitamin D metabolism, skeletal formation and  
 CC mineralization. Phosphatonins are used to treat hyperphosphatemia, renal  
 CC osteodystrophy, secondary hyperparathyroidism, osteitis fibrosa cystica  
 CC or gout. It is used to prepare a medicament for treating X-linked  
 CC hypophosphatemic rickets, hereditary hypophosphatemic rickets with  
 CC hypocalcemia (HHRH), hypomineralized bone lesions, stunted growth in  
 CC juveniles, oncogenic hypophosphatemic osteomalacia, renal phosphate  
 CC leakage, renal osteodystrophy, osteoporosis, vitamin-D resistant rickets,  
 CC end organ resistance, renal Fanconi syndrome, autosomal rickets, Paget's  
 CC disease, kidney failure, renal tubular acidosis, cystic fibrosis or  
 CC sprue. Phosphatonin polynucleotides are useful as molecular weight  
 CC markers on Southern gels, as diagnostic probes for detecting the presence  
 CC of a specific mRNA. Phosphatonin polypeptides are also useful for  
 CC identifying agonists and antagonists, compounds which bind to  
 CC phosphatonin and drug candidates for therapy of phosphate metabolism  
 CC disorders. The present sequence represents a truncated form of  
 CC phosphatonin (MEPP).

CC  
 XX  
 SQ Sequence 430 AA:

Query Match 100.0%; Score 39; DB 22; Length 430;  
 Best Local Similarity 50.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 12  
 1:1:1:1:1:1  
 DB 154 DNDISPFSGDQ 165

RESULT 14  
 AAB82922  
 ID AAB82922 standard; Protein: 509 AA.

XX AAB82922;

XX 21-DEC-2001 (first entry)

DE Human osteoregulin (mature polypeptide).

KW Osteoregulin; human; bone; homeostasis; adipose; calcification;  
 KW atherosclerosis; osteoporosis; osteopathic; antiarteriosclerotic;  
 KW therapy.

OS Homo sapiens.

XX EPI130098-A2.

XX 05-SEP-2001.

XX 27-FEB-2001; 2001EP-0301768.

XX 29-FEB-2000; 2000US-185617P.

PR 22-SEP-2000; 2000US-234500P.

XX (PFI2 ) PFIZER PROD INC.

PI Brown TA, De Wet JR, Gowen LC, Hames LM;

XX WPI: 2001-604111/69.

DR N-PSDB: AAH26810.

PT Novel osteoregulin polypeptide useful for regulating bone homeostasis,  
 PT adiposity and calcification of atherosclerotic plaques comprises  
 PT measuring the activity of osteoregulin -

XX Claim 1; Page 54-55; 90pp; English.

XX The present sequence is that of human osteoregulin mature  
 CC polypeptide, i.e. lacking an N-terminal signal sequence.  
 CC Osteoregulin is a novel protein which plays a role in regulating  
 CC bone homeostasis, adiposity, and the calcification of  
 CC atherosclerotic plaques. 2 splice variants of human osteoregulin  
 CC were identified (see also AAB82923). The invention provides novel

CC osteoregulin proteins, nucleic acids which encode them, vectors,  
 CC antibodies, host cells which express heterologous osteoregulins, and  
 CC animal cells and mammals with a targeted disruption of an  
 CC osteoregulin gene. The invention also provides screening assays  
 CC to identify modulators of osteoregulin activity as well as methods  
 CC of treating mammals for diseases or disorders associated with  
 CC osteoregulin activity. The modulators of activity may be useful  
 CC in the manufacture of a medicament for, as well as for treating, a  
 CC mammal in need of regulation of bone mass and/or density, adiposity,  
 CC vascular flexibility, and/or atherosclerotic plaque calcification  
 CC (claimed), for treating and preventing osteoporosis, and for  
 CC stimulating bone repair and regeneration.

CC  
 XX  
 SQ Sequence 509 AA:

Query Match 100.0%; Score 39; DB 22; Length 509;  
 Best Local Similarity 50.0%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 12  
 1:1:1:1:1:1  
 DB 233 DNDISPFSGDQ 244

RESULT 15  
 AAB82920  
 ID AAB82920 standard; Protein: 525 AA.

XX AAB82920;

XX 21-DEC-2001 (first entry)

DE Human osteoregulin.

KW Osteoregulin; human; bone; homeostasis; adipose; calcification;  
 KW atherosclerosis; osteoporosis; osteopathic; antiarteriosclerotic;  
 KW therapy.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..37 /Label= Signal\_peptide

FT Protein 38..525 /Label= Mature\_protein

XX EPI130098-A2.

XX 05-SEP-2001.

XX 27-FEB-2001; 2001EP-0301768.

XX 29-FEB-2000; 2000US-185617P.

PR 22-SEP-2000; 2000US-234500P.

XX (PFI2 ) PFIZER PROD INC.

PI Brown TA, De Wet JR, Gowen LC, Hames LM;

XX WPI: 2001-604111/69.

DR N-PSDB: AAH26808.

PT Novel osteoregulin polypeptide useful for regulating bone homeostasis,  
 PT adiposity and calcification of atherosclerotic plaques comprises  
 PT measuring the activity of osteoregulin -

XX Claim 1; Page 45-47; 90pp; English.

XX The present sequence is that of human osteoregulin, a novel protein  
 CC which plays a role in regulating bone homeostasis, adiposity, and  
 CC the calcification of atherosclerotic plaques. The sequence is  
 CC predicted from the nucleotide sequence (see AAH26808) of isolated  
 CC osteoblast cDNA. A splice variant of human osteoregulin was also



CC identified (see AAB82921). The invention provides novel  
CC osteoregulin proteins, nucleic acids which encode them, vectors, and  
CC antibodies, host cells which express heterologous osteoregulins, and  
CC animal cells and mammals with a targeted disruption of an  
CC osteoregulin gene. The invention also provides screening assays  
CC to identify modulators of osteoregulin activity as well as methods  
CC of treating mammals for diseases or disorders associated with  
CC osteoregulin activity. The modulators of activity may be useful  
CC in the manufacture of a medicament for, as well as for treating, a  
CC mammal in need of regulation of bone mass and/or density, adiposity,  
CC vascular flexibility, and/or atherosclerotic plaque calcification  
CC (claimed), for treating and preventing osteoporosis, and for  
CC stimulating bone repair and regeneration.

XX  
SQ Sequence 525 AA:

Query Match 100.0%; Score 39; DB 22; Length 525;  
Best Local Similarity 50.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 DDXSXFXGXXQ 12  
|:|:|:|:|:|:  
Db 249 DNDISFSGDQ 260

Search completed: March 18, 2003, 08:29:50  
Job time : 33.5 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 08:29:30 ; Search time 11.5 Seconds  
(without alignments)  
30.702 Million cell updates/sec

Title: US-09-812-485a-43  
Perfect score: 39  
Sequence: 1 DDXSXFXGXQ 12

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	92.3	14	4	US-09-242-435-17
2	36	92.3	1391	4	US-09-106-568E-8
3	35	89.7	294	4	US-09-077-955-29
4	35	89.7	534	4	US-09-124-541-1
5	35	89.7	633	4	US-09-134-001C-3162
6	34	87.2	159	3	US-08-992-176-6
7	34	87.2	188	3	US-09-130-663-30
8	34	87.2	188	3	US-09-332-934-14
9	34	87.2	188	4	US-09-432-335-30
10	34	87.2	188	4	US-09-614-022-30
11	33	84.6	221	1	US-07-754-918A-6
12	33	84.6	1442	2	US-08-316-650-12
13	33	84.6	1442	5	PCT-US85-02251-12
14	33	84.6	2206	1	US-07-852-260-2
15	33	84.6	2206	2	US-08-461-503-2
16	33	84.6	2206	4	US-08-465-250-2
17	32	82.1	12	5	PCT-US93-05701-4
18	32	82.1	261	4	US-09-329-418-6
19	32	82.1	261	4	US-09-531-914-6
20	32	82.1	315	2	US-08-910-927B-3
21	32	82.1	315	4	US-09-270-270-3
22	32	82.1	420	3	US-09-329-418-8
23	32	82.1	420	4	US-09-531-914-8
24	32	82.1	518	3	US-09-329-418-3
25	32	82.1	518	3	US-09-329-418-4
26	32	82.1	518	3	US-09-329-418-5
27	32	82.1	518	3	US-09-329-418-9

28	32	82.1	518	4	US-09-531-914-3	Sequence 3, Appl1
29	32	82.1	518	4	US-09-531-914-4	Sequence 4, Appl1
30	32	82.1	518	4	US-09-531-914-5	Sequence 5, Appl1
31	32	82.1	518	4	US-09-531-914-9	Sequence 9, Appl1
32	32	82.1	833	4	US-09-514-302-3	Sequence 3, Appl1
33	32	82.1	1938	4	US-09-514-302-2	Sequence 2, Appl1
34	31	79.5	20	1	US-08-440-861-35	Sequence 35, Appl
35	31	79.5	31	3	US-08-926-842B-58	Sequence 58, Appl
36	31	79.5	70	4	US-08-847-065-16	Sequence 16, Appl
37	31	79.5	177	1	US-08-446-920-11	Sequence 11, Appl
38	31	79.5	240	4	US-09-355-166-19	Sequence 19, Appl
39	31	79.5	248	4	US-09-032-523-1	Sequence 1, Appl1
40	31	79.5	263	1	US-07-971-096-4	Sequence 4, Appl1
41	31	79.5	263	1	US-08-175-096-4	Sequence 4, Appl1
42	31	79.5	263	4	US-08-413-974-6	Sequence 6, Appl1
43	31	79.5	263	4	US-08-434-418-6	Sequence 6, Appl1
44	31	79.5	263	4	US-08-433-288-6	Sequence 6, Appl1
45	31	79.5	263	4	US-08-174-739A-6	Sequence 6, Appl1

#### ALIGNMENTS

```
RESULT 1
US-09-242-435-17
; Sequence 17, Application US/09242435
; Patent No. 6461863
; GENERAL INFORMATION:
; APPLICANT: JARVIS, DONALD L.
; TITLE OF INVENTION: MODIFYING INSECT CELL GLYCOSYLATION PATHWAYS WITH
; FILE REFERENCE: UWYO:00205
; CURRENT APPLICATION NUMBER: US/09/242,435
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-242-435-17

Query Match      92.3%; Score 36; DB 4; Length 14;
Best Local Similarity 41.7%; Pred. No. 2.7;
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY      1 DDXSXFXGXQ 12
DB      1 DDDPSRFSKAE 12

RESULT 2
US-09-106-568E-8
; Sequence 8, Application US/09106568E
; Patent No. 6455248
; GENERAL INFORMATION:
; APPLICANT: Bhattacharjee, J.
; APPLICANT: Suvarna, Kalavathi
; APPLICANT: Bhattacharjee, Vaakeer
; TITLE OF INVENTION: METHODS AND REAGENTS FOR DETECTING FUNGAL PATHOGENS IN
; TITLE OF INVENTION: A BIOLOGICAL SAMPLE
; FILE REFERENCE: 96/247-A
; CURRENT APPLICATION NUMBER: US/09/106,568E
; CURRENT FILING DATE: 1998-06-29
; PRIOR APPLICATION NUMBER: 08/650,809
; PRIOR FILING DATE: 1997-05-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: Microsoft Word 97
; SEQ ID NO 8
; LENGTH: 1391
```

TYPE: PRT  
ORGANISM: Candida albicans  
US-09-106-568E-8

Query Match  
Best Local Similarity 41.7%; Score 36; DB 4; Length 1391;  
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 12  
1:1:1:1:1:1  
DB 221 DDDSGYRGAIQ 232

RESULT 3  
US-09-077-955-29  
Sequence 29, Application US/09077955A.  
Patent No. 6413740

GENERAL INFORMATION:  
APPLICANT: Valenzuela et al., David M.  
TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS  
FILE REFERENCE: REG195-B-PCT-US  
CURRENT APPLICATION NUMBER: US/09/077,955A  
CURRENT FILING DATE: 1998-09-10  
EARLIER APPLICATION NUMBER: PCT/US96/20696  
EARLIER FILING DATE: 1996-12-13  
EARLIER APPLICATION NUMBER: 08/644,271  
EARLIER FILING DATE: 1996-05-10  
EARLIER APPLICATION NUMBER: 60/008,657  
EARLIER FILING DATE: 1995-12-15  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 29  
LENGTH: 294  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-077-955-29

Query Match  
Best Local Similarity 41.7%; Score 35; DB 4; Length 294;  
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 12  
1:1:1:1:1:1  
DB 19 DDDKGFDGAIQ 30

RESULT 4

US-09-124-541-1  
Sequence 1, Application US/09124541A  
Patent No. 6229066  
GENERAL INFORMATION:  
APPLICANT: Morris Ph.D., Roy O.  
TITLE OF INVENTION: A CYTOKININ OXIDASE  
FILE REFERENCE: UMO1490  
CURRENT APPLICATION NUMBER: US/09/124,541A  
CURRENT FILING DATE: 1998-07-29  
EARLIER APPLICATION NUMBER: 60/054,268  
EARLIER FILING DATE: 1997-07-30  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 1  
LENGTH: 534  
TYPE: PRT  
ORGANISM: Zea mays  
US-09-124-541-1

Query Match  
Best Local Similarity 41.7%; Score 35; DB 4; Length 534;  
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 12  
1:1:1:1:1:1  
DB 409 DFDGCVFKGILQ 420

RESULT 5  
US-09-134-001C-3162  
Sequence 3162, Application US/09134001C  
Patent No. 6380370

GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 3162  
LENGTH: 633  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3162

Query Match  
Best Local Similarity 41.7%; Score 35; DB 4; Length 633;  
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 12  
1:1:1:1:1:1  
DB 334 DHDEGFGICQ 345

RESULT 6  
US-08-992-176-6  
Sequence 6, Application US/08992176  
Patent No. 6125331

GENERAL INFORMATION:  
APPLICANT: TOH, Hiroyuki  
TITLE OF INVENTION: STRUCTURAL ALIGNMENT METHOD MAKING USE OF A DOUBLE  
FILE REFERENCE: 9200-0001-2  
CURRENT APPLICATION NUMBER: US/08/992,176  
CURRENT FILING DATE: 1997-12-17  
EARLIER APPLICATION NUMBER: JP 8-340727  
EARLIER FILING DATE: 1996-12-20  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 6  
LENGTH: 159  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: beta-protein  
US-08-992-176-6

Query Match  
Best Local Similarity 45.5%; Score 34; DB 3; Length 159;  
Matches 5; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 11  
1:1:1:1:1:1  
DB 3 DFDISKFLGFW 13

RESULT 7  
US-09-130-663-30  
Sequence 30, Application US/09130663A  
Patent No. 6020163

GENERAL INFORMATION:  
APPLICANT: Conklin, Darrell C.  
TITLE OF INVENTION: LIPOCALIN HOMOLOG

```

: FILE REFERENCE: 97-24
: CURRENT APPLICATION NUMBER: US/09/130,663A
: CURRENT FILING DATE: 1998-08-05
: EARLIER APPLICATION NUMBER: 60/054,867
: EARLIER FILING DATE: 1997-08-06
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 30
: LENGTH: 188
: TYPE: PR1
: ORGANISM: Rattus norvegicus
: OS-09-130-663-30

```

Query Match	87.28;	Score 34;	DB 3;	Length 188;
Best Local Similarity	45.58;	Pred. No. 1.3e+02;		
Matches	5;	Conservative	6;	Mismatches 0;
				Indels

```
QY 1 DDXDSXFXGXX 11  
1:1:1:1:1:1:  
Db 27 DFDISKFLGFW 37
```

```

1  RESULT 8
2  US-09-332-934-14
3  Sequence 14, Application US/09332934
4  Patent No. 611413
5  GENERAL INFORMATION:
6  APPLICANT: Murry, Lynn, E.
7  APPLICANT: Tang, Tom, Y.
8  APPLICANT: Baughn, Mariah, R.
9  TITLE OF INVENTION: LIPOCALIN FAMILY PROTEIN
10 FILE REFERENCE: PC-0005 US
11 CURRENT APPLICATION NUMBER: US/09/332,934
12 CURRENT FILING DATE: 1999-06-14
13 NUMBER OF SEQ ID NOS: 14
14 SOFTWARE: PERL Program
15 SEQ ID NO 14
16 LENGTH: 188
17 TYPE: PRT
18 ORGANISM: Rattus norvegicus
19 FEATURE:
20 NAME/KEY:
21 OTHER INFORMATION: g113824
22 PUBLICATION INFORMATION:
23 US-09-332-934-14

```

Query Match	87.2%;	Score 34;	DB 3;	Length 188;
Best Local Similarity	45.5%;	Pred. No. 1.3e+02;		
Matches	5;	Conservative	6;	Mismatches 0;
				Indels

```
QY      1 DDXDSXFXGXX 11
        1:1:1:1:1::
Db      27 DFDISKFLGFW 37
```

```

1  RESULT 9
2  US-09-432-335-30
3  : Sequence 30 Application US/09432335
4  : Patent No. 6143720
5  : GENERAL INFORMATION:
6  : APPLICANT: Conklin, Darrell C.
7  : TITLE OF INVENTION: LIPOCALIN HOMOLOG
8  : FILE REFERENCE: 97-24
9  : CURRENT APPLICATION NUMBER: US/09/432,335
10 : CURRENT FILING DATE: 1999-11-02
11 : EARLIER APPLICATION NUMBER: 09/130,663
12 : EARLIER FILING DATE: 1998-08-06
13 : EARLIER APPLICATION NUMBER: 60/054,867
14 : EARLIER FILING DATE: 1997-08-06
15 : NUMBER OF SEQ ID NOS: 30
16 : SOFTWARE: FastSeq For Windows Version 3.0.
17 : SEQ ID NO 30
18 : LENGTH: 188

```

```

; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-432-335-30

```

Query Match	87.28;	Score 34;	DB 4;	Length 188;
Best Local Similarity	45.58;	Pred. No. 1.3e+02;		
Matches	5;	Conservative	6;	Mismatches 0;
				Indels 0;
				Gaps 0;

```
QY      1 DXDXSXFXGXX 11
          |::|::|::
Db      27 DFDISKFLGFW 37
```

```

00000000 RESULT 0
00000001 US-09-614-022-30
00000002 ; Sequence 30, Application US/09614022
00000003 ; Patent No. 6365716
00000004 ;
00000005 ; GENERAL INFORMATION:
00000006 ; APPLICANT: Conklin, Darrell C.
00000007 ; TITLE OF INVENTION: LIPOCALIN HOMOLOG
00000008 ; FILE REFERENCE: 97-24
00000009 ; CURRENT APPLICATION NUMBER: US/09/614,022
00000010 ; CURRENT FILING DATE: 2000-07-11
00000011 ;
00000012 ; PRIOR APPLICATION NUMBER: 09/130,663
00000013 ; PRIOR FILING DATE: 1998-08-06
00000014 ; PRIOR APPLICATION NUMBER: 60/054,867
00000015 ; PRIOR FILING DATE: 1997-08-06
00000016 ; NUMBER OF SEQ ID NOS: 30
00000017 ; SOFTWARE: FastSeq for Windows Version 3.0
00000018 ; SEQ ID NO 30
00000019 ;
00000020 ; LENGTH: 188
00000021 ;
00000022 ; TYPE: PRT
00000023 ;
00000024 ; ORGANISM: Rattus norvegicus
00000025 ;
00000026 US-09-614-022-30

```

Query Match 87.2%; Score 34; DB 4; Length 188;  
Best Local Similarity 45.5%; Pred. No. 1.3e+02;  
Matches 5; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 DXDXSXFXX 11
    1:1:1:1:1:1:
Db 27 DFDISKFLGFW 37
```

RESULT 11  
US-07-754-918A-6  
Sequence 6, Application US/07754918A  
Patent No. 5286484  
GENERAL INFORMATION:  
APPLICANT: Rodriguez, R.S. et al  
TITLE OF INVENTION: NUCLEOTIDE SEQUENCE CODING FOR AN  
OUTER MEMBRANE PROTEIN FROM NEISSERIA MENINGITIDIS AND USE  
TITLE OF INVENTION: OF SAID PROTEIN IN VACCINE PREPARATIONS  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Stanger, Michaelson, Spivak and Wallace, Esq.  
STREET: Parkway 109 Office Center, 328 Newman Springs  
STREET: Road, P. O. Box 8489  
CITY: Red Bank  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5 1/4" 360kb IBM compatible diskette  
COMPUTER: IBM PS/2 Model 80  
OPERATING SYSTEM: MS-DOS 5.0  
SOFTWARE: Microsoft Word 5.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/754,918A  
FILING DATE: 19910905  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Michaelson, Peter L.

REGISTRATION NUMBER: 30090  
REFERENCE/DOCKET NUMBER: Centro-2R  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908)530-6671  
TELEFAX: (908)530-6584  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 221 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Sequence which includes variable regions of  
US-07-754-918A-6

Query Match 84.6%; Score 33; DB 1; Length 221;  
Best Local Similarity 45.5%; Pred. No. 2.5e+02;  
Matches 5; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 2 XDXXFXGXQ 12  
DB 130 PDFFSGGSVQ 140

RESULT 12  
US-08-316-650-12  
Sequence 12, Application US/08316650  
Patent No. 5942496  
GENERAL INFORMATION:  
APPLICANT: Bonadio, Jeffrey  
APPLICANT: Roessler, Blake J.  
APPLICANT: Goldstein, Steven A.  
APPLICANT: Lin, Wushan  
TITLE OF INVENTION: METHODS AND COMPOSITIONS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/316,650  
FILING DATE: 30-SEP-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/199,780  
FILING DATE: 30-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, David L.  
REGISTRATION NUMBER: 32,165  
REFERENCE/DOCKET NUMBER: UMIC:008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1442 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-316-650-12

Query Match 84.6%; Score 33; DB 2; Length 1442;

Best Local Similarity 45.5%; Pred. No. 2.3e+03;  
Matches 5; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 2 XDXXFXGXQ 12  
DB 1173 IDMSAFAGLQ 1183

RESULT 13  
PCT-US95-02251-12  
Sequence 12, Application PC/TUS9502251  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
SOFTWARE: Patent Release #1.0, Version  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/02251  
FILING DATE: CONCURRENTLY HERewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/316,650  
FILING DATE: 30-SEP-1994  
CLASSIFICATION:  
APPLICATION NUMBER: US 08/199,780  
FILING DATE: 18-FEB-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, David L.  
REGISTRATION NUMBER: 32,165  
REFERENCE/DOCKET NUMBER: UMIC009P--  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1442 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US95-02251-12

Query Match 84.6%; Score 33; DB 5; Length 1442;  
Best Local Similarity 45.5%; Pred. No. 2.3e+03;  
Matches 5; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 2 XDXXFXGXQ 12  
DB 1173 IDMSAFAGLQ 1183

RESULT 14  
US-07-852-260-2  
Sequence 2, Application US/07852260  
Patent No. 5525715  
GENERAL INFORMATION:  
APPLICANT: Racanelli, Vincent  
APPLICANT: Talem, Joanne M.

```
; APPLICANT: Weeks-Levy, Carolyn L.
; TITLE OF INVENTION: METHODS FOR PRODUCING RNA VIRUSES FROM
; TITLE OF INVENTION: CDNA
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/852,260
; FILING DATE: 19920619
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 36607-B-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEO ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2206 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-852-260-2

Query Match      84.6%; Score 33; DB 1; Length 2206;
Best Local Similarity 45.5%; Pred. No. 3.8e+03;
Matches 5; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY      2 XDXXFXGXQ 12
Db      1283 PDPHFQGYKQ 1293

RESULT 15
US-08-461-503-2
; Sequence 2, Application US/08461503
; Patent No. 5834302
; GENERAL INFORMATION:
; APPLICANT: Racanelli, Vincent
; APPLICANT: Talem, Joanne M.
; TITLE OF INVENTION: METHODS FOR PRODUCING RNA VIRUSES
; TITLE OF INVENTION: FROM CDNA
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,503
; FILING DATE: 5-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 36607-D-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEO ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2206 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-461-503-2

Query Match      84.6%; Score 33; DB 2; Length 2206;
Best Local Similarity 45.5%; Pred. No. 3.8e+03;
Matches 5; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY      2 XDXXFXGXQ 12
Db      1283 PDPHFQGYKQ 1293

Search completed: March 18, 2003, 08:32:22
Job time : 12.5 secs
```

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 08:29:54 ; Search time 11 Seconds  
(without alignments)  
50.282 Million cell updates/sec

Title: US-09-812-485a-42

Perfect score: 66

Sequence: 1 DNDISPFSGDGQ 12

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published.Applications\_AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	66	100.0	12	9	US-09-812-485a-42
2	66	100.0	15	9	US-09-812-485a-47
3	66	100.0	23	9	US-09-812-485a-49
4	66	100.0	40	9	US-09-812-485a-23
5	66	100.0	97	9	US-09-812-485a-1
6	66	100.0	525	10	US-09-812-485a-2
7	60	90.9	15	9	US-09-812-485a-48
8	60	90.9	33	9	US-09-812-485a-30
9	55	83.3	19	10	US-09-812-485a-6
10	49	74.2	30	9	US-09-812-485a-26
11	48.5	73.5	35	9	US-09-812-485a-25
12	48.5	73.5	40	9	US-09-812-485a-24
13	41	62.1	2039	9	US-10-192-584-7
14	41	62.1	2042	9	US-10-192-584-6
15	39	59.1	15	9	US-09-812-485a-46
16	39	59.1	347	10	US-09-815-242-11959
17	37	56.1	354	10	US-09-825-414-70
18	37	56.1	623	9	US-10-108-605-125
19	37	56.1	623	9	US-10-108-605-129

20	37	56.1	1198	9	US-09-975-719-405	Sequence 405, App
21	36	54.5	413	10	US-09-746-491-43	Sequence 43, Appl
22	36	54.5	437	10	US-09-815-242-11837	Sequence 11837, A
23	36	54.5	623	10	US-09-815-242-11837	Sequence 11837, A
24	36	54.5	877	10	US-09-746-491-47	Sequence 47, Appl
25	36	54.5	906	10	US-09-905-983-46	Sequence 46, Appl
26	36	54.5	906	10	US-09-746-491-46	Sequence 46, Appl
27	36	54.5	906	10	US-09-746-491-48	Sequence 48, Appl
28	36	54.5	912	10	US-09-905-983-2	Sequence 2, Appl
29	36	54.5	912	10	US-09-746-491-49	Sequence 49, Appl
30	35	53.0	114	10	US-09-864-761-36859	Sequence 36859, A
31	35	53.0	188	10	US-09-815-242-4981	Sequence 4981, Ap
32	35	53.0	188	10	US-09-815-242-10949	Sequence 10949, A
33	35	53.0	205	10	US-09-841-132-564	Sequence 564, App
34	35	53.0	240	10	US-09-950-368-19	Sequence 19, Appl
35	35	53.0	326	9	US-09-738-626-4959	Sequence 4959, Ap
36	35	53.0	450	10	US-09-815-242-13497	Sequence 13497, A
37	35	53.0	532	10	US-09-891-160-2	Sequence 2, Appl
38	35	53.0	607	10	US-09-734-002-2	Sequence 2, Appl
39	35	53.0	607	10	US-09-801-196-29	Sequence 29, Appl
40	34	51.5	17	10	US-09-861-294-33	Sequence 33, Appl
41	34	51.5	58	10	US-09-864-761-40191	Sequence 40191, A
42	34	51.5	147	10	US-09-797-481-4	Sequence 4, Appl
43	34	51.5	147	10	US-09-844-736-6	Sequence 6, Appl
44	34	51.5	182	10	US-09-784-810A-29	Sequence 29, Appl
45	34	51.5	290	10	US-09-925-297-817	Sequence 817, App

## ALIGNMENTS

RESULT 1  
US-09-812-485a-42  
; Sequence 42, Application US/09812485A  
; Publication No. US20020197267A1  
; GENERAL INFORMATION:  
; APPLICANT: Kumagai, Yoshihara  
; APPLICANT: Blacher, Russel  
; TITLE OF INVENTION: Integrin Binding Motif Containing  
; TITLE OF INVENTION: Peptides and Methods of Treating Skeletal Diseases  
; FILE REFERENCE: BEAR-006CIP  
; CURRENT APPLICATION NUMBER: US/09/812,485A  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 09/641,034  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 42  
; LENGTH: 12  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: calcium binding motif  
US-09-812-485a-42  
Query Match 100.0%; Score 66; DB 9; Length 12;  
Best local Similarity 100.0%; Pred. No. 3.1e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DNDISPFSGDGQ 12  
Db 1 DNDISPFSGDGQ 12  
RESULT 2  
US-09-812-485a-47  
; Sequence 47, Application US/09812485A  
; Publication No. US20020197267A1  
; GENERAL INFORMATION:  
; APPLICANT: Kumagai, Yoshihara  
; APPLICANT: Blacher, Russel  
; APPLICANT: Yoneda, Toshiyuki

```
; TITLE OF INVENTION: Integrin Binding Motif Containing Skeletal Diseases
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; CURRENT FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: D-00004 peptide
; NAME/KEY: AMIDATION
; LOCATION: 15
US-09-812-485a-47
```

```
Query Match          100.0%; Score 66; DB 9; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DNDISPFSGDQ 12
    |||
Db 4 DNDISPFSGDQ 15
```

```
RESULT 3
; Sequence 49, Application US/09812485A
; Publication No. US20020197267A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; APPLICANT: Yoneda, Toshiyuki
; TITLE OF INVENTION: Integrin Binding Motif Containing Skeletal Diseases
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; CURRENT FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: D-00006 peptide
; NAME/KEY: AMIDATION
; LOCATION: 15
US-09-812-485a-49
```

```
Query Match          100.0%; Score 66; DB 9; Length 23;
Best Local Similarity 100.0%; Pred. No. 6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DNDISPFSGDQ 12
    |||
Db 8 DNDISPFSGDQ 19
```

```
RESULT 4
; Sequence 23, Application US/09812485A
; Publication No. US20020197267A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; APPLICANT: Yoneda, Toshiyuki
; TITLE OF INVENTION: Integrin Binding Motif Containing
```

```
; TITLE OF INVENTION: Peptides and Methods of Treating Skeletal Diseases
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; CURRENT FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: peptidic compound
US-09-812-485a-23
```

```
Query Match          100.0%; Score 66; DB 9; Length 40;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DNDISPFSGDQ 12
    |||
Db 3 DNDISPFSGDQ 14
```

```
RESULT 5
; Sequence 1, Application US/09812485A
; Publication No. US20020197267A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; APPLICANT: Yoneda, Toshiyuki
; TITLE OF INVENTION: Integrin Binding Motif Containing Skeletal Diseases
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; CURRENT FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: peptidic compound
US-09-812-485a-1
```

```
Query Match          100.0%; Score 66; DB 9; Length 97;
Best Local Similarity 100.0%; Pred. No. 0.00026;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DNDISPFSGDQ 12
    |||
Db 50 DNDISPFSGDQ 61
```

```
RESULT 6
; Sequence 6, Application US/09814550
; Patent No. US20020102641A1
; GENERAL INFORMATION:
; APPLICANT: Schiavi, Susan
; APPLICANT: Maden, Stephen
; APPLICANT: Manavalan, Parthasarathy
; APPLICANT: Levine, Michael
; APPLICANT: Jan de Beur, Suzanne
; TITLE OF INVENTION: ONCOGENIC OSTEOMALACIA-RELATED GENE 1
; FILE REFERENCE: 5014US
; CURRENT APPLICATION NUMBER: US/09/814,550
; CURRENT FILING DATE: 2001-03-22
```

```

; PRIOR APPLICATION NUMBER: US 60/191,786
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 60/241,598
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-814-550-2
```

```

Query Match          100.0%; Score 66; DB 10; Length 525;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DNDISPFSGDQ 12
        |||
Db       249 DNDISPFSGDQ 260
```

```

RESULT 7
; Sequence 48, Application US/09812485A
; Publication No. US20020197267A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; TITLE OF INVENTION: Integrin Binding Motif Containing
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: D-00005 peptide
; NAME/KEY: AMIDATION
; LOCATION: 15
US-09-812-485A-48
```

```

Query Match          90.9%; Score 60; DB 9; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00039;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 NDISPFSGDQ 12
        |||
Db       1 NDISPFSGDQ 11
```

```

RESULT 8
; Sequence 30, Application US/09812485A
; Publication No. US20020197267A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; TITLE OF INVENTION: Integrin Binding Motif Containing
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; NUMBER OF SEQ ID NOS: 50
; TYPE: PRT
```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: peptidic compound
US-09-812-485A-30
```

```

Query Match          90.9%; Score 60; DB 9; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.00088;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 NDISPFSGDQ 12
        |||
Db       1 NDISPFSGDQ 11
```

```

RESULT 9
; Sequence 6, Application US/09814550
; Patent No. US20020102641A1
; GENERAL INFORMATION:
; APPLICANT: Schiavi, Susan
; APPLICANT: Madden, Stephen
; APPLICANT: Manavalan, Parthasarathy
; APPLICANT: Levine, Michael
; APPLICANT: Jan de Beur, Suzanne
; TITLE OF INVENTION: ONCOGENIC OSTEOBLASTIC-RELATED GENE 1
; FILE REFERENCE: 5014US
; CURRENT APPLICATION NUMBER: US/09/814,550
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,786
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 60/241,598
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-814-550-6
```

```

Query Match          83.3%; Score 55; DB 10; Length 19;
Best Local Similarity 100.0%; Pred. No. 0.0034;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DNDISPFSGD 10
        |||
Db       10 DNDISPFSGD 19
```

```

RESULT 10
; Sequence 26, Application US/09812485A
; Publication No. US20020197267A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; TITLE OF INVENTION: Integrin Binding Motif Containing
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 30
; TYPE: PRT
```

ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: peptidic compound  
US-09-812-485a-26

Query Match  
Best Local Similarity 74.2%; Score 49; DB 9; Length 30;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 NDISPFGD 10  
Db 1 NDISPFGD 9

RESULT 11  
US-09-812-485a-25  
Sequence 25, Application US/09812485A  
Publication No. US20020197267A1  
GENERAL INFORMATION:  
APPLICANT: Kumagai, Yoshinari  
APPLICANT: Blacher, Russel  
APPLICANT: Yoneda, Toshiyuki  
TITLE OF INVENTION: Integrin Binding Motif Containing  
Peptides and Methods of Treating Skeletal Diseases  
FILE REFERENCE: BEAR-006CIP  
CURRENT APPLICATION NUMBER: US/09/812,485A  
CURRENT FILING DATE: 2001-03-19  
PRIORITY APPLICATION NUMBER: 09/641,034  
PRIOR FILING DATE: 2000-08-16  
NUMBER OF SEQ ID NOS: 50  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 25  
LENGTH: 35  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: peptidic compound  
US-09-812-485a-25

Query Match  
Best Local Similarity 73.5%; Score 48.5; DB 9; Length 35;  
Matches 11; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 2 NDISPF--SGDQ 12  
Db 1 NDISPFGDSDGQ 14

RESULT 12  
US-09-812-485a-24  
Sequence 24, Application US/09812485A  
Publication No. US20020197267A1  
GENERAL INFORMATION:  
APPLICANT: Kumagai, Yoshinari  
APPLICANT: Blacher, Russel  
APPLICANT: Yoneda, Toshiyuki  
TITLE OF INVENTION: Integrin Binding Motif Containing  
Peptides and Methods of Treating Skeletal Diseases  
FILE REFERENCE: BEAR-006CIP  
CURRENT APPLICATION NUMBER: US/09/812,485A  
CURRENT FILING DATE: 2001-03-19  
PRIORITY APPLICATION NUMBER: 09/641,034  
PRIOR FILING DATE: 2000-08-16  
NUMBER OF SEQ ID NOS: 50  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 24  
LENGTH: 40  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: peptidic compound  
US-09-812-485a-24

Query Match  
Best Local Similarity 73.5%; Score 48.5; DB 9; Length 40;  
Matches 11; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 2 NDI---SPFGSDGQ 12  
Db 1 NDIRGDSFGSDGQ 14

RESULT 13  
US-10-192-584-7  
Sequence 7, Application US/10192584  
Publication No. US20030027987A1  
GENERAL INFORMATION:  
APPLICANT: TOKUNAGA, Ei-ji  
SAKAGUCHI, Masashi  
MATSUO, Kazuo  
HAMADA, Fukuaburo  
TOKITOSHI, Sachio  
TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS  
PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEWMARK  
STREET: 624 Ninth Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/192,584  
FILING DATE: 11-Jul-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/077,098  
FILING DATE: 19-May-1998  
APPLICATION NUMBER: PCT/JP97/03222  
FILING DATE: 12-SEP-1997  
APPLICATION NUMBER: JP 27,148/1996  
FILING DATE: 19-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: KORNBAU, Anne M.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: TOKUNAGA-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2039 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Query Match  
Best Local Similarity 62.1%; Score 41; DB 9; Length 2039;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 DNDISPFGSD 10  
Db 1599 DGDISPFGSD 1608

RESULT 14  
US-10-192-584-6  
Sequence 6, Application US/10192584

```
; Publication No. US20030027967A1
;
; GENERAL INFORMATION:
; APPLICANT: TOKUNAGA, Eiichi
;             SAKAGUCHI, Masashi
;             MATSUO, Kazuo
;             HAMADA, Fukusaburo
;             TOKIYOSHI, Sachio
; TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS
;             PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME
;
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 624 Ninth Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 11-Jul-2002
; APPLICATION NUMBER: US/10/192,584
; FILING DATE: 11-Jul-2002
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/09/077,098
; FILING DATE: 19-May-1998
; APPLICATION NUMBER: PCT/JP97/03222
; FILING DATE: 12-SEP-1997
; APPLICATION NUMBER: JP 27,148/1996
; FILING DATE: 19-SEP-1996
;
; ATTORNEY/AGENT INFORMATION:
; NAME: KORNBAU, Anne M.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TOKUNAGA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
;
; INFORMATION FOR SEQ ID NO: 6:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2042 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
;
; US-10-192-584-6
;
; Query Match          62.1%; Score 41; DB 9; Length 2042;
; Best Local Similarity 80.0%; Pred. No. 93;
; Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
;
; QY 1 DNDISPGSD 10
;      | | | | |
; Db 1591 DGDISTSGD 1600
;
; RESULT 15
; US-09-812-485A-46
; Sequence 46, Application US/09812485A
; Publication No. US20020197267A1
;
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; APPLICANT: Yoneda, Toshiyuki
; TITLE OF INVENTION: Integritin Binding Motif Containing
;             Peptides and Methods of Treating Skeletal Diseases
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; CURRENT FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; PRIOR FILING DATE: 2000-08-16
```

```
; NUMBER OF SEQ ID NOS: 50
;
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: AMIDATION
; LOCATION: 15
; US-09-812-485A-46
;
; Query Match          59.1%; Score 39; DB 9; Length 15;
; Best Local Similarity 100.0%; Pred. No. 1.3;
; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 DNDISPF 7
;      | | | | |
; Db 9 DNDISPF 15
;
; Search completed: March 18, 2003, 08:32:50
; Job time : 12 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 08:29:54 ; Search time 11 Seconds  
(without alignments)  
50.282 Million cell updates/sec

Title: US-09-812-485a-43

Perfect score: 39

Sequence: 1 DXDXSFXGXGXQ 12

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications-AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	100.0	12	9 US-09-812-485a-42	Sequence 42, Appl
2	39	100.0	12	9 US-09-812-485a-43	Sequence 43, Appl
3	39	100.0	15	9 US-09-812-485a-47	Sequence 47, Appl
4	39	100.0	23	9 US-09-812-485a-49	Sequence 49, Appl
5	39	100.0	40	9 US-09-812-485a-23	Sequence 23, Appl
6	39	100.0	97	9 US-09-812-485a-1	Sequence 1, Appl
7	39	100.0	525	10 US-09-814-550-2	Sequence 2, Appl
8	36	92.3	504	10 US-09-764-864-1173	Sequence 1173, Ap
9	36	92.3	1004	10 US-09-767-215-2	Sequence 2, Appl
10	36	92.3	1391	9 US-09-994-595-8	Sequence 8, Appl
11	35	89.7	294	9 US-10-016-283-29	Sequence 29, Appl
12	35	89.7	384	10 US-09-886-055-87	Sequence 87, Appl
13	35	89.7	450	10 US-09-815-242-5214	Sequence 5214, Ap
14	35	89.7	536	9 US-09-738-626-6061	Sequence 6061, Ap
15	35	89.7	611	10 US-09-829-549A-48	Sequence 48, Appl
16	35	89.7	1017	10 US-09-897-056-5	Sequence 5, Appl
17	34	87.2	131	9 US-09-712-363-191	Sequence 191, Appl
18	34	87.2	139	9 US-10-032-159A-16	Sequence 16, Appl
19	34	87.2	421	9 US-09-738-626-5227	Sequence 5227, Ap

20	34	87.2	432	10	US-09-815-242-10274	Sequence 10274, A
21	34	87.2	436	10	US-09-815-242-13866	Sequence 13866, A
22	34	87.2	436	10	US-09-815-242-11154	Sequence 11154, A
23	34	87.2	1138	10	US-09-767-215-5	Sequence 5, Appl
24	34	87.2	1325	10	US-09-741-669-304	Sequence 304, Appl
25	33	84.6	15	9	US-09-812-485a-48	Sequence 48, Appl
26	33	84.6	19	10	US-09-814-550-6	Sequence 6, Appl
27	33	84.6	33	9	US-09-812-485a-30	Sequence 30, Appl
28	33	84.6	59	10	US-09-864-761-39774	Sequence 39774, A
29	33	84.6	66	10	US-09-864-761-46305	Sequence 46305, A
30	33	84.6	270	10	US-09-911-826A-8	Sequence 8, Appl
31	33	84.6	484	9	US-09-738-626-6755	Sequence 6755, Appl
32	33	84.6	500	9	US-09-738-626-4296	Sequence 4296, Ap
33	33	84.6	500	9	US-09-738-626-6393	Sequence 6393, Ap
34	33	84.6	722	10	US-09-815-242-10796	Sequence 10796, A
35	32	82.1	142	10	US-09-862-027-31	Sequence 31, Appl
36	32	82.1	233	10	US-09-925-300-1529	Sequence 1529, Ap
37	32	82.1	311	9	US-09-738-626-3862	Sequence 3862, Ap
38	32	82.1	315	9	US-09-984-245-200	Sequence 200, Appl
39	32	82.1	315	10	US-09-847-809A-3	Sequence 3, Appl
40	32	82.1	478	9	US-09-866-050A-512	Sequence 512, Appl
41	32	82.1	497	10	US-09-862-027-8	Sequence 8, Appl
42	32	82.1	518	10	US-09-771-161A-231	Sequence 231, Appl
43	32	82.1	655	9	US-09-738-626-5440	Sequence 5440, Ap
44	32	82.1	727	10	US-09-822-246-2	Sequence 2, Appl
45	32	82.1	833	9	US-10-014-436-3	Sequence 3, Appl

#### ALIGNMENTS

RESULT 1  
US-09-812-485a-42  
; Sequence 42, Application US/09812485A  
; Publication No. US20020197267A1  
; GENERAL INFORMATION:  
; APPLICANT: Kumagai, Yoshinari  
; APPLICANT: Blacher, Russel  
; APPLICANT: Yoneda, Toshiyuki  
; TITLE OF INVENTION: Integrin Binding Motif Containing Peptides and Methods of Treating Skeletal Diseases  
; FILE REFERENCE: BEAR-006CIP  
; CURRENT APPLICATION NUMBER: US/09/812,485A  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 09/641,034  
; PRIOR FILING DATE: 2000-08-16  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 42  
; LENGTH: 12  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: calcium binding motif  
US-09-812-485a-42  
  
Query Match 100.0%; Score 39; DB 9; Length 12;  
Best Local Similarity 50.0%; Pred.No. 0.9;  
Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 DXDXSFXGXGXQ 12  
|:|:|:|:|:|:  
DB 1 DNDISFSGGQ 12  
  
RESULT 2  
US-09-812-485a-43  
; Sequence 43, Application US/09812485A  
; Publication No. US20020197267A1  
; GENERAL INFORMATION:  
; APPLICANT: Kumagai, Yoshinari  
; APPLICANT: Blacher, Russel  
; APPLICANT: Yoneda, Toshiyuki

```
; TITLE OF INVENTION: Integrin Binding Motif Containing
; TITLE OF INVENTION: Peptides and Methods of Treating Skeletal Diseases
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: calcium-binding motif
; NAME/KEY: VARIANT
; LOCATION: 2, 4, 6, 8, 10, 11
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-812-485A-43
```

```
Query Match          100.0%; Score 39; DB 9; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.9;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DDXSXFXGXQ 12
|:|:|:|:|:|
Db 1 DDXSXFXGXQ 12
```

```
RESULT 3
US-09-812-485A-47
; Sequence 47, Application US/09812485A
; Publication No. US20020197267A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; APPLICANT: Yoneda, Toshiyuki
; TITLE OF INVENTION: Integrin Binding Motif Containing
; TITLE OF INVENTION: Peptides and Methods of Treating Skeletal Diseases
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; CURRENT FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: D-00004 peptide
; NAME/KEY: AMIDATION
; LOCATION: 15
US-09-812-485A-47
```

```
Query Match          100.0%; Score 39; DB 9; Length 15;
Best Local Similarity 50.0%; Pred. No. 1.2;
Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DDXSXFXGXQ 12
|:|:|:|:|:|
Db 4 DNDISFSGDQ 15
```

```
RESULT 4
US-09-812-485A-49
; Sequence 49, Application US/09812485A
; Publication No. US20020197267A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; APPLICANT: Yoneda, Toshiyuki
```

```
; TITLE OF INVENTION: Integrin Binding Motif Containing
; TITLE OF INVENTION: Peptides and Methods of Treating Skeletal Diseases
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: D-00006 peptide
; NAME/KEY: AMIDATION
; LOCATION: 15
US-09-812-485A-49
```

```
Query Match          100.0%; Score 39; DB 9; Length 23;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DDXSXFXGXQ 12
|:|:|:|:|:|
Db 8 DNDISFSGDQ 19
```

```
RESULT 5
US-09-812-485A-23
; Sequence 23, Application US/09812485A
; Publication No. US20020197267A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; APPLICANT: Yoneda, Toshiyuki
; TITLE OF INVENTION: Integrin Binding Motif Containing
; TITLE OF INVENTION: Peptides and Methods of Treating Skeletal Diseases
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; CURRENT FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: peptidic compound
US-09-812-485A-23
```

```
Query Match          100.0%; Score 39; DB 9; Length 40;
Best Local Similarity 50.0%; Pred. No. 3.5;
Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DDXSXFXGXQ 12
|:|:|:|:|:|
Db 3 DNDISFSGDQ 14
```

```
RESULT 6
US-09-812-485A-1
; Sequence 1, Application US/09812485A
; Publication No. US20020197267A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; APPLICANT: Yoneda, Toshiyuki
; TITLE OF INVENTION: Integrin Binding Motif Containing
; TITLE OF INVENTION: Peptides and Methods of Treating Skeletal Diseases
; FILE REFERENCE: BEAR-006CIP
```



```
; CURRENT APPLICATION NUMBER: US/09/812,485A
; CURRENT FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: peptidic compound
US-09-812-485A-1
```

```
Query Match          100.0%; Score 39; DB 9; Length 97;
Best Local Similarity 50.0%; Pred. No. 9.5;
Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DDXSXFXGXGXQ 12
        |:|:|:|:|:|
Db      50 DNDISFSGDGQ 61
```

```
RESULT 7
US-09-814-550-2
; Sequence 2, Application US/09814550
; Patent No. US20020102641A1
; GENERAL INFORMATION:
; APPLICANT: Schlavi, Susan
; APPLICANT: Madden, Stephen
; APPLICANT: Manavalan, Parthasarathy
; APPLICANT: Levine, Michael
; APPLICANT: Jan de Beur, Suzanne
; TITLE OF INVENTION: ONCOGENIC OSTEOCLAST-RELATED GENE 1
; FILE REFERENCE: 5014US
; CURRENT APPLICATION NUMBER: US/09/814,550
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,786
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 60/241,598
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-814-550-2
```

```
Query Match          100.0%; Score 39; DB 10; Length 525;
Best Local Similarity 50.0%; Pred. No. 64;
Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DDXSXFXGXGXQ 12
        |:|:|:|:|:|
Db      249 DNDISFSGDGQ 260
```

```
RESULT 8
US-09-764-864-1173
; Sequence 1173, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1173
; LENGTH: 504
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (91)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (92)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (167)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (198)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (216)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (486)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1173
```

```
Query Match          92.3%; Score 36; DB 10; Length 504;
Best Local Similarity 41.7%; Pred. No. 2.1e+02;
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DDXSXFXGXGXQ 12
        |:|:|:|:|:|
Db      359 DVDLSQFQGLME 370
```

```
RESULT 9
US-09-767-215-2
; Sequence 2, Application US/09767215
; Patent No. US20020081636A1
; GENERAL INFORMATION:
; APPLICANT: Berlin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; FILE REFERENCE: 07334-142001
; CURRENT APPLICATION NUMBER: US/09/767,215
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: 60/181,159
; PRIOR FILING DATE: 2000-02-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1004
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-767-215-2
```

```
Query Match          92.3%; Score 36; DB 10; Length 1004;
Best Local Similarity 41.7%; Pred. No. 4.5e+02;
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DDXSXFXGXGXQ 12
        |:|:|:|:|:|
Db      109 DVDLSNFQGLME 120
```

```
RESULT 10
US-09-994-595-8
; Sequence 8, Application US/0994595
; Publication No. US20030039981A1
; GENERAL INFORMATION:
; APPLICANT: Bhattacharjee, J.
; APPLICANT: Suvarna, Kalavati
; APPLICANT: Bhattacharjee, Vasker
; TITLE OF INVENTION: METHODS AND REAGENTS FOR DETECTING FUNGAL PATHOGENS IN
; FILE REFERENCE: 96,247-A
```

```
; CURRENT APPLICATION NUMBER: US/09/994,595
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 08/650,809
; PRIOR FILING DATE: 1997-05-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: Microsoft Word 97
; SEQ ID NO: 8
; LENGTH: 1391
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-994-595-8

Query Match
Best Local Similarity 92.3%; Score 36; DB 9; Length 1391;
41.7%; Pred. No. 6.5e+02;
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDXSXFXGXQ 12
|:|:|:|:|:|
Db 221 DDDMSGRGAIQ 232

RESULT 11
US-10-016-283-29
; Sequence 29, Application US/10016283
; Patent No. US20020164702A1
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/10/016,283
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US/09/077,955A
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: PCT/US96/20696
; PRIOR FILING DATE: 1996-12-13
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 29
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-016-283-29

Query Match
Best Local Similarity 89.7%; Score 35; DB 9; Length 294;
41.7%; Pred. No. 1.7e+02;
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 DDXSXFXGXQ 12
|:|:|:|:|:|
Db 19 DDDKGFDAIQ 30

RESULT 12
US-09-886-055-87
; Sequence 87, Application US/0986055
; Patent No. US20020132273A1
; GENERAL INFORMATION:
; APPLICANT: STRYER, LOBERT
; APPLICANT: ZOZULYA, SERGEY
; TITLE OF INVENTION: RECEPTOR FINGERPRINTING, SENSORY PERCEPTION, AND
FILE REFERENCE: 07803-0277150
; CURRENT APPLICATION NUMBER: US/09/886,055
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: 60/213,812
; PRIOR FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 522
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 87
; LENGTH: 384
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; NAME/KEY: MOD_RBS
; LOCATION: (68)
; OTHER INFORMATION: Variable amino acid
US-09-886-055-87

Query Match
Best Local Similarity 89.7%; Score 35; DB 10; Length 384;
41.7%; Pred. No. 2.3e+02;
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 DDXSXFXGXQ 12
|:|:|:|:|:|
Db 22 DSDFKNFICIQ 33

RESULT 13
US-09-815-242-5214
; Sequence 5214, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Mall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 5214
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5214

Query Match
Best Local Similarity 89.7%; Score 35; DB 10; Length 450;
41.7%; Pred. No. 2.7e+02;
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 DDXSXFXGXQ 12
|:|:|:|:|:|
Db 82 DDDPGFPGKIQ 93

RESULT 14
US-09-738-626-6061
; Sequence 6061, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHITAI, KEIKO
```

```
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, MAOKO
; APPLICANT: SENO, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6061
; LENGTH: 536
; TYPE: PRF
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6061
```

```
Query Match          89.7%; Score 35; DB 9; Length 536;
Best Local Similarity 41.7%; Pred. No. 3.3e+02;
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 DDXSXFXGXQ 12
    :|:|:|:|:|
Db 315 EMDLSEFVGDIQ 326
```

```
RESULT 15
US-09-829-549A-48
; Sequence 48, Application US/09829549A
; Patent No. US20020052484A1
; GENERAL INFORMATION:
; APPLICANT: The Curators of the University of Missouri
; TITLE OF INVENTION: PHAGE DISPLAY SELECTION OF ANTI FUNGAL PEPTIDES
; FILE REFERENCE: DMO 1521.1
; CURRENT APPLICATION NUMBER: US/09/829,549A
; CURRENT FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: US 60/195,785
; PRIOR FILING DATE: 2000-04-10
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 48
; LENGTH: 611
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(85)
; OTHER INFORMATION: Mat-alpha secretory sequence
; NAME/KEY: DOMAIN
; LOCATION: (86)..(600)
; OTHER INFORMATION: Cytokinin oxidase 1
; NAME/KEY: DOMAIN
; LOCATION: (601)..(602)
; OTHER INFORMATION: Linker
; NAME/KEY: DOMAIN
; LOCATION: (603)..(611)
; OTHER INFORMATION: Random peptide Pc 87
; US-09-829-549A-48
```

```
Query Match          89.7%; Score 35; DB 10; Length 611;
Best Local Similarity 41.7%; Pred. No. 3.8e+02;
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 DDXSXFXGXQ 12
    :|:|:|:|:|
Db 475 DFDGRGVFGILQ 486
```

**THIS PAGE BLANK (USPTO)**



C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
C:Accession: t37881  
R:Skellon, J.; Churche, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, August 1997  
A:Reference number: 221751  
A:Accession: t37881  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-144 <SKE>  
A:Cross-references: EMBL:298597; PIDN:CAB11225.1; GSPDB:GN00066; SPDB:SPAC17H9.16  
A:Experimental source: strain 972h-; cosmid c17H9  
C:Genetics:  
A:Gene: SPDB:SPAC17H9.16  
A:Map position: 1  
A:Introns: 131/3  
C:Superfamily: mitochondrial receptor complex chain MOM22  
C:Keywords: mitochondrion

Query Match 92.3%; Score 36; DB 2; Length 144;  
Best Local Similarity 41.7%; Pred. No. 41;  
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 12  
|:|:|:|:|:|:  
Db 37 DSDSDFEGLER 48

RESULT 3  
EB4466  
hypothetical protein At2g05260 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: EB4466  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanKen, S.E.; Umayam, L.; Tallon, L.;  
Nuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: EB4466  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-358 <STO>  
A:Cross-references: GB:AE002093; NID:g4755196; PIDN:AD29063.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g05260  
A:Map position: 2

Query Match 92.3%; Score 36; DB 2; Length 358;  
Best Local Similarity 41.7%; Pred. No. 1.1e+02;  
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 12  
|:|:|:|:|:|:  
Db 74 DADSFIFGIFE 85

RESULT 4  
S57264  
L-aminoadipate-semialdehyde dehydrogenase (EC 1.2.1.31) - fission yeast (Schizosaccharom  
C:Species: Schizosaccharomyces pombe  
C>Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 17-Mar-1999  
C:Accession: S57264  
R:Ford, R.A.; Bhattacharjee, J.K.  
Curr. Genet. 28, 131-137, 1995  
A:Title: Molecular properties of the lys1(+) gene and the regulation of alpha-aminoadipa  
A:Reference number: S57264; MUID:96020159; PMID:8590464  
A:Accession: S57264  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-470 <FOR>  
A:Cross-references: EMBL:U15923  
C:Keywords: oxidoreductase

Query Match 92.3%; Score 36; DB 2; Length 470;  
Best Local Similarity 41.7%; Pred. No. 1.5e+02;  
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 12  
|:|:|:|:|:|:  
Db 270 DDDWSGRCATQ 281

RESULT 5  
T01882  
hypothetical protein F8M12.9 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 24-Mar-1999  
C:Accession: T01882  
R:Madsen, C.; Graves, T.; Cotton, M.; Modde, T.  
submitted to the EMBL Data Library, July 1998  
A:Description: The sequence of A. thaliana F8M12.  
A:Reference number: Z14450  
A:Accession: T01882  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-649 <MAD>  
A:Cross-references: EMBL:AF080118; NID:g3513725; PID:g3513738  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 4  
A:Introns: 6/2; 201/1; 291/3; 372/1; 410/3; 490/3; 576/1  
A>Note: F8M12.9

Query Match 92.3%; Score 36; DB 2; Length 649;  
Best Local Similarity 41.7%; Pred. No. 2.1e+02;  
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 12  
|:|:|:|:|:|:  
Db 60 DADNSIFGIFE 71

RESULT 6  
T04290  
hypothetical protein F25124.160 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 30-Apr-1999  
C:Accession: T04290  
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro  
submitted to the Protein Sequence Database, March 1999  
A:Reference number: Z15261  
A:Accession: T04290  
A:Molecule type: DNA  
A:Residues: 1-665 <BEV>  
A:Cross-references: EMBL:AL049525  
A:Experimental source: cultivar Columbia; BAC clone F25124  
C:Genetics:  
A:Map position: 4  
A:Introns: 22/2; 218/2; 307/3; 388/1; 426/3; 506/3; 592/1  
A>Note: F25124.160

Query Match 92.3%; Score 36; DB 2; Length 665;  
Best Local Similarity 41.7%; Pred. No. 2.1e+02;  
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 12  
|:|:|:|:|:|:  
Db 76 DADNSIFGIFE 87

RESULT 7  
T01757  
hypothetical protein A\_IG002P16.20 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 22-Oct-1999

```

C:Accession: T01757
R:Miller, N.; Beck, C.; Kramer, J.
submitted to the EMBL Data Library, June 1997
A:Description: The sequence of A. thaliana IG002P16.
A:Reference number: 214421
A:Accession: T01757
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1133 <MIL>
A:Cross-references: EMBL:AF007270; NID:g2191157; PID:g2191168; GSPDB:GN00063; ATSP:A_IG002P16.20
A:Map position: 5
A:Introns: 56/3; 92/3; 246/3; 288/3; 334/2; 532/1; 599/1; 641/1; 868/1; 934/3; 1022/3

Query Match
Best Local Similarity 41.7%; Score 36; DB 2; Length 1133;
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDXSXFXGXQ 12
|:|:|:|:|:|:|
Db 918 DRDQSMFGPPE 929

RESULT 8
T50973
probable alpha-aminoadipate reductase large subunit [imported] - Neurospora crassa
N:Alternate names: protein B24P7.280
C:Species: Neurospora crassa
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 17-Nov-2000
R:Schulle, U.; Allyn, V.; Hohelsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, July 2000
A:Reference number: 225286
A:Accession: T50973
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1174 <SCH>
A:Cross-references: EMBL:AL389890; GSPDB:GN00116; NCSP:B24P7.280
A:Experimental source: BAC clone B24P7; strain OR7A4
C:Genetics:
A:Gene: NCSP:B24P7.280
A:Map position: 6
A:Introns: 96/3
C:Superfamily: acetate-CoA ligase homology; acyl carrier protein homology
C:Keywords: carrier protein
F:71-610/Domain: acetate-CoA ligase homology <ACL>
F:634-705/Domain: acyl carrier protein homology <ACP>

Query Match
Best Local Similarity 41.7%; Score 36; DB 2; Length 1174;
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

PY 1 DDXSXFXGXQ 12
|:|:|:|:|:|:|
Db 11 DLDMSGYVGSIO 22

RESULT 9
A90413
conserved hypothetical protein [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001
C:Accession: A90413
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.; Barrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to Genbank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: A90413
A:Status: preliminary
A:Molecule type: DNA

```

A:Residues: 1-138 <KUD>  
A:Cross-references: GB:AE006641; NID:913815718; PIDN:AAK42560.1; GSPDB:GN00155  
C:Genetics:  
A:Gene: SS02412  
C:Superfamily: hypothetical protein M21272

Query Match 89.7%, Score 35; DB 2; Length 138;  
Best Local Similarity 41.7%; Pred. No. 61;  
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDXSXFXGXQ 12  
|:|:|:|:|:  
Db 76 DLDXSRFEGPFR 87

RESULT 10

A82691  
tRNA pseudouridine synthase A XF1373 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
C:Accession: A82691  
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq  
Nature 406, 151-157, 2000  
A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82615; MUID:20365711; PMID:10910347  
A>Note: for a complete list of authors see reference number A59328 below  
A:Accession: A82691  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-257 <STM>  
A:Cross-references: GB:AE003968; GB:AE003849; NID:99106363; PIDN:AAF84182.1; GSPDB:GN  
A:Experimental source: strain 9a5c  
R:Stimpson, A.J.G.; Reinach, F.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.  
Bioness, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir  
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
Submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; FR  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitaajima, J.P.; Krieger, J.E.; Kurame, E.E.; La  
chado, M.A.; Medeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.;  
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Saw  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv  
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF1373  
C:Superfamily: tRNA-pseudouridine synthase I

Query Match 89.7%, Score 35; DB 2; Length 257;  
Best Local Similarity 41.7%; Pred. No. 1.2e+02;  
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDXSXFXGXQ 12  
|:|:|:|:|:  
Db 9 EYDSSEFLGMCQ 20

RESULT 11

C69832  
methyl-accepting chemotaxis protein homolog yhfV - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
C:Accession: C69832  
R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Be  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.R.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gal  
teck, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, J  
Koster, P.; Koningsstedt, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinc  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; May

Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelli, Rieger, M.; Rivolta, C.; Roeha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schoeter, R.; Scioffone, F.; Sekiguchi, J.; Sekowska, A.; Serot, T.; Miners, P.; Wipat, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, A.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A:Reference number: A69580; MUID:98044033; PMID:9384377  
 A:Accession: C69832  
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-432 <KUN>  
 A:Cross-references: GB:299109; GB:AL009126; NID:92633260; PIDN:CAB12878.1; PID:92633374  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: ynfV  
 C:Superfamily: Pyrococcus horikoshii probable methyl-accepting chemotaxis protein; trans

Query Match 89.7%; Score 35; DB 2; Length 432;  
 Best Local Similarity 41.7%; Pred. No. 2.1e+02;  
 Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 12  
 |:|: |:|:|:|  
 Db 399 DVDLQAFUGLQ 410

RESULT 12  
 E82973  
 Conserved hypothetical protein PA5393 [Imported] - *Pseudomonas aeruginosa* (strain PA01)  
 C:Species: *Pseudomonas aeruginosa*  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: E82973  
 R:Stover, C.K.; Plam, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lathig, K.; Lim, J.; Lory, S.; Olson, M.V.  
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.  
 A:Reference number: A82950; MUID:20437337; PMID:10984043  
 A:Accession: E82973  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-450 <STO>  
 A:Cross-references: GB:AE004951; GB:AE004091; NID:99951705; PIDN:AA08778.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA5393

Query Match 89.7%; Score 35; DB 2; Length 450;  
 Best Local Similarity 41.7%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 12  
 |:|: |:|:|:|  
 Db 82 DDDPGFPGKQ 93

RESULT 13  
 T21623  
 Hypothetical protein F32A7.4 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T21623  
 R:Kershaw, J.  
 A:Title: Submitted to the EMBL Data Library, November 1996  
 A:Reference number: Z19451  
 A:Accession: T21623  
 A:Status: Preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-529 <WIL>  
 A:Cross-references: EMBL:Z83107; PIDN:CAB05500.1; GSPDB:GN00019; CESP:F32A7.4  
 A:Experimental source: clone F32A7  
 C:Genetics:

A:Gene: CESP:F32A7.4  
 A:Map position: 1  
 A:Introns: 27/3; 60/2; 250/3; 280/3; 436/2; 502/3

Query Match 89.7%; Score 35; DB 2; Length 529;  
 Best Local Similarity 41.7%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 12  
 |:|: |:|:|:|  
 Db 131 DFDKESFGKQV 142

RESULT 14  
 T01500  
 cytochrome oxidase 1 - maize  
 C:Species: *Zea mays* (maize)  
 C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 29-Oct-1999  
 C:Accession: T01500  
 R:Morris, R.O.; Laskey, J.G.  
 A:Title: Submitted to the EMBL Data Library, January 1998  
 A:Description: A glycosylated cytochrome oxidase from maize.  
 A:Reference number: Z14336  
 A:Accession: T01500  
 A:Status: Preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-534 <MOR>  
 A:Cross-references: EMBL:AF044603; NID:93341977; PIDN:AAC27500.1; PID:93341978  
 C:Genetics:  
 A:Gene: CKX1  
 A:Introns: 206/1; 437/2

Query Match 89.7%; Score 35; DB 2; Length 534;  
 Best Local Similarity 41.7%; Pred. No. 2.7e+02;  
 Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 12  
 |:|: |:|:|:|  
 Db 409 DEDRGVFGKILQ 420

RESULT 15  
 T51929  
 cytochrome oxidase [Imported] - maize  
 C:Species: *Zea mays* (maize)  
 C:Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 20-Oct-2000  
 C:Accession: T51929  
 R:Houba-Herlin, N.; Pethé, C.; D'Alayer, J.; Laloue, M.  
 A:Title: Cytochrome oxidase from *Zea mays*: purification, cDNA cloning and expression  
 A:Reference number: Z25869  
 A:Accession: T51929  
 A:Status: Preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-534 <HOU>  
 A:Cross-references: EMBL:Y18377; PIDN:CAA77151.1  
 A:Experimental source: cultivar nobilis  
 C:Genetics:  
 A:Gene: cko

Query Match 89.7%; Score 35; DB 2; Length 534;  
 Best Local Similarity 41.7%; Pred. No. 2.7e+02;  
 Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 12  
 |:|: |:|:|:|  
 Db 409 DEDRGVFGKILQ 420

Search completed: March 18, 2003, 08:31:53  
 Job time : 15.5 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 08:27:42 ; Search time 7.5 Seconds

(without alignments)  
66.362 Million cell updates/sec

Title: US-09-812-485a-43

Perfect score: 39  
Sequence: 1 DXDXSFXGXQ 12

Scoring table: BLOSUM62X  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	100.0	233	1 CLC1_YEAST	P17891 saccharomyc
2	36	92.3	622	1 YA41_HUMAN	Q9UPW0 homo sapien
3	36	92.3	999	1 CARE_MOUSE	Q9KX16 mus musculu
4	36	92.3	1004	1 CARE_HUMAN	Q12572 candida alb
5	36	92.3	1391	1 LYS2_CANAL	P40976 schizosacch
6	36	92.3	1419	1 LYS2_SCHPO	007621 bacillus su
7	35	89.7	432	1 HMA7_BACSU	Q91066 oryza sativ
8	35	89.7	532	1 CKX1_ORISA	Q91066 oryza sativ
9	35	89.7	534	1 CKX1_MAIZE	Q91066 oryza sativ
10	35	89.7	1021	1 MANA_RHOMR	P49425 rhodothermu
11	34	87.2	54	1 ELHB_APLCA	P11925 aplysia cal
12	34	87.2	54	1 ELHB_APLCA	P11925 aplysia cal
13	34	87.2	133	1 Y843_AOUAE	Q67011 aquilex aeo
14	34	87.2	173	1 ELHA_APLCA	P01360 aplysia cal
15	34	87.2	179	1 ELIN_AGR75	Q57259 agrobacteri
16	34	87.2	188	1 ERBP_RAT	P06911 rattus norv
17	34	87.2	189	1 ICYA_MANSE	P00305 manduca sex
18	34	87.2	206	1 ICYA_MANSE	Q00630 manduca sex
19	34	87.2	303	1 LPXC_PSEAE	P47205 pseudomonas
20	34	87.2	332	1 Y713_CHILM	Q09173 chlamydia m
21	34	87.2	341	1 GUN1_CRYFL	Q04469 cryptococcu
22	34	87.2	411	1 GLYA_CIOAB	Q979V1 clostridium
23	34	87.2	424	1 MS11_ARATH	Q22467 arabidopsis
24	34	87.2	424	1 MS11_ARATH	Q22467 arabidopsis
25	34	87.2	431	1 ENO_ECOLI	P08324 escherichia
26	34	87.2	431	1 ENO_ECOLI	Q8X926 salmonella
27	34	87.2	431	1 ENO_SALTY	Q8X926 salmonella
28	34	87.2	431	1 ENO_YERPE	Q8X926 yersinia pe
29	34	87.2	434	1 ENO_VIRCH	Q9KPS5 vibrio chol
30	34	87.2	435	1 ENO_HAERI	Q57391 gallus gall
31	34	87.2	474	1 HOSM_PENCH	P43806 haemophilus
32	34	87.2	503	1 PUR1_PASMT	Q91688 pasteurella
33	34	87.2	531	1 NUON_MYCTU	Q53308 mycobacteri

34	34	87.2	531	1	YQCG_BACSU	P45942	bacillus su
35	34	87.2	569	1	Y139_MYCPN	P75497	mycoplasma
36	34	87.2	637	1	VART_TREAC	Q9H611	thermoplasma
37	34	87.2	1325	1	YDEK_ECOLI	P32051	escherichia
38	34	87.2	3829	1	SACS_HUMAN	Q9N2J4	homo sapien
39	33	84.6	142	1	OSMC_ECOLI	P23929	escherichia
40	33	84.6	330	1	ODBA_BACSU	P37940	bacillus su
41	33	84.6	342	1	UF01_GAEEL	Q19584	caenorhabdi
42	33	84.6	362	1	ERD1_YEAST	P16151	saccharomyc
43	33	84.6	369	1	CAL2_CHICK	P02460	gallus gall
44	33	84.6	387	1	OMA2_NEIMC	P18194	neisseria m
45	33	84.6	402	1	TH1B_CANTR	Q04677	candida tro

## ALIGNMENTS

```

RESULT 1
ID CLC1_YEAST          STANDARD;          PRT;          233 AA.
AC P17891;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Clathrin light chain (CLC).
GN CLC1 OR YGRI67M.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-39.
RX MEDLINE=91009480; PubMed=2211819.
RA Silveira L.A., Wong D.H., Maslarz F.R., Shekman R.;
RT "Yeast Clathrin has a distinctive light chain that is important for
RT cell growth."
RL J. Cell Biol. 111:1437-1449(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=5288c;
RX MEDLINE=97435481; PubMed=9290212;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT chromosome VII."
RL Yeast 13:1077-1090(1997).
RN [3]
RP SEQUENCE OF 10-233 FROM N.A.
RA Hebling U., Hofmann B., Dellus H.;
RT Submitting (MAY-1996) to the EMBL/GenBank/DDJJ databases.
RL
CC -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYEDRAL COAT OF
CC COATED PITTS & VESICLES. IN YEAST, IT IS INVOLVED IN THE RETENTION
CC OF PROTEINS IN AN INTRACELLULAR MEMBRANE COMPARTMENT, PRESUMABLY
CC THE TRANS-GOLGI. THE YEAST LIGHT CHAIN IS IMPORTANT FOR CELL
CC GROWTH. THE LIGHT CHAIN MAY HELP TO PROPERLY ORIENT THE ASSEMBLY/
CC DISASSEMBLY OF THE CLATHRIN COATS.
CC -!- SUBUNIT: CLATHRIN COATS ARE FORMED FROM MOLECULES CONTAINING 3
CC HEAVY CHAINS AND 3 LIGHT CHAINS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITTS AND
CC VESICLES.
CC -!- MISCELLANEOUS: CLC1 BINDS CALCIUM, AND CALMODULIN IN PRESENCE OF
CC CALCIUM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X52272; CAA36515.1; -
CC DR EMBL: 272953; CAA97193.1; -
CC DR EMBL: 272953; CAA97193.1; -

```

```

DR PIR: A36425; A36425.
DR SGD: S0003399; CLC1.
DR InterPro: IPR000996; Clathrin_1g_ch.
DR Pfam: PF01086; Clathrin_1g_ch. 1.
DR PROSITE: PS00224; CLATHRIN_LIGHT_CHAIN_1; FALSE-NEG.
DR Coated pits; Calcium-binding; Calmodulin-binding; Coiled coil.
KW BINDING 144 204 HEAVY CHAIN (POTENTIAL).
FT SIMILAR 144 204 TO REGIONS OF INTERMEDIATE FILAMENTS AND
FT OTHER PROTEINS THAT FORM COILED-COIL
FT STRUCTURES.
FT SIMILAR 193 213 TO THE BRAIN SPECIFIC INSERT IN LCA/LCB.
SQ SEQUENCE 233 AA; 26531 MW; 278B815780ECB3 CRC64;

Query Match
Best Local Similarity 50.0%; Score 39; DB 1; Length 233;
Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDXSXFXGXQXQ 12
1:1:1:1:1:1
DB 97 DDEFSTFEKANQ 108

RESULT 2
YA41_HUMAN STANDARD; PRT; 622 AA.
AC Q9UPW0;
ID YA41_HUMAN STANDARD; PRT; 622 AA.
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative forkhead box protein KIAA1041.
GN KIAA1041.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RC MEDLINE=99397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 6:197-205(1999).
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC -----
CC EMBL: AB028964; BAA82993.1; -.
CC HSSP: 063245; 2HPH.
CC TRANSFAC: T04179; -.
CC InterPro: IPR001766; TF_Fork_head.
CC Pfam: PF00250; Fork_head. 1.
CC PRINTS: PR00053; FORKHEAD.
CC PRODOM: PD000425; TF_Fork_head. 1.
CC SMART: SM00339; FH; 1.
CC PROSITE: PS00657; FORK_HEAD_1; 1.
CC PROSITE: PS00658; FORK_HEAD_2; 1.
CC PROSITE: PS00039; FORK_HEAD_3; 1.
KW Hypothetical protein; Transcription regulation; DNA-binding;
KW Nuclear protein.
FT DNA_BIND 77 168 FORK-HEAD.
FT SEQUENCE 622 AA; 68928 MW; F8C577D71A7980FF CRC64;

```

```

Query Match
Best Local Similarity 41.7%; Score 36; DB 1; Length 622;
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDXSXFXGXQXQ 12
1:1:1:1:1:1
DB 477 DVDLSQFQGLME 488

RESULT 3
CARE_MOUSE STANDARD; PRT; 999 AA.
ID CARE_MOUSE STANDARD; PRT; 999 AA.
AC Q99KF0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Caspase recruitment domain protein 14 (Bcl10-interacting MAGUK protein
DE 2) (Bim2).
GN CARD14 OR BIM2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21391892; PubMed=11387339;
RA McAllister-Lucas L.M., Inohara N., Lucas P.C., Ruland J., Benito A.,
RA Li Q., Chen S., Chen F.F., Yamaoka S., Verma I.M., Mak T.W.,
RA Nunez G.;
RT "Bim1, a MAGUK family member linking protein kinase C activation to
RT Bcl10-mediated NF-kappa B induction."
RL J. Biol. Chem. 276:30589-30597(2001).
RN [2]
RP SEQUENCE OF 82-743 FROM N.A.
RC TISSUE=Breast;
RA Strausberg R.;
RT Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: Activates NF-kappa B via Bcl10 and IKK.
CC -1- SUBUNIT: CARD14 and Bcl10 bind to each other by CARD-CARD
CC interaction (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -1- CAUTION: Supposed to contain a SH3 domain which is not detected by
CC PROSITE, Pfam or SMART.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC -----
CC EMBL: AF363457; AAK60137.1; -.
CC EMBL: BC004692; AAH04692.1; -.
CC InterPro: IPR000619; Guanylate_kin.
CC DR InterPro: IPR001478; PDZ.
CC Pfam: PF00595; PDZ; 1.
CC PROSITE: PS02029; CARD; 1.
CC PROSITE: PS00856; GUANYLATE_KINASE_1; FALSE-NEG.
CC PROSITE: PS00852; GUANYLATE_KINASE_2; 1.
CC PROSITE: PS0106; PDZ; 1.
KW Coiled coil.
FT DOMAIN 15 107 CARD.
FT DOMAIN 125 411 COILED COIL (POTENTIAL).
FT DOMAIN 572 655 PDZ.
FT DOMAIN 854 986 GUANYLATE KINASE.
FT CONFLICT 736 743 OAOOOLLA -> HLEHDRS (IN REF. 2).
SQ SEQUENCE 999 AA; 113496 MW; D18350DA12430255 CRC64;

Query Match
Score 36; DB 1; Length 999;

```

[illegible]

```

DR InterPro:IPR001478; PDZ.
DR SMART: SM00072; GUKC; 1.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS50209; CARD; 1.
DR PROSITE: PS50856; GUANYLATE_KINASE_1; FALSE_NG.
DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE: PS50106; PDZ; 1.
KW Coiled coil.
FT DOMAIN 15 107 CARD.
FT DOMAIN 128 409 COILED COIL (POTENTIAL).
FT DOMAIN 568 658 PDZ.
FT DOMAIN 858 990 GUANYLATE_KINASE.
FT CONFLICT 619 671 DYRSEPLFRAVLEDTTLEEAVGLLRVDGECCLASYVNTD
FT GYKRLQDLLEAK -> SRARPLSPGLMGTVAGGVQAD
FT FTSPRCRSTLGMVASALSMADVKRSAML (IN REF. 2;
FT AAH01326).
SQ SEQUENCE 1004 AA; 113299 MW; 7469BBB56BE0673 CRC64;

Query Match 92.3%; Score 36; DB 1; Length 1004;
Best Local Similarity 41.7%; Pred. No. 1.le+02;
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDXSXFXGXQX 12
Db 109 DVDFSNFSGIME 120

RESULT 5
LYS2_CANAL
ID LYS2_CANAL STANDARD; PRT; 1391 AA.
AC Q12572;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Aminoacidate-semialdehyde dehydrogenase large subunit (EC 1.2.1.31)
DE (Alpha-aminoacidate reductase) (Alpha-AK).
GN LYS2.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11651 / B792;
RX MEDLINE=9828267; Pubmed=9560434;
RA Suvanna K., Seah L., Bhattacharjee V., Bhattacharjee J.K.;
RT "Molecular analysis of the LYS2 gene of Candida albicans: homology to
RT peptide antibiotic synthetases and the regulation of the alpha-
RT aminoacidate reductase."
RT Curr. Genet. 33:268-275(1998).
- FUNCTION: CATALYZES THE ACTIVATION OF ALPHA-AMINOACIDATE BY
- ATP-DEPENDENT ADENYLATION AND THE REDUCTION OF ACTIVATED
- ALPHA-AMINOACIDATE BY NADPH.
CC CATALYTIC ACTIVITY: L-2-aminoacidate 6-semialdehyde + NADP(+) +
CC H(2)O = L-2-aminoacidate + NADPH.
CC COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE
CC (POTENTIAL).
CC PATHWAY: Lysine biosynthesis; sixth step.
CC SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT (BY
CC SIMILARITY).
CC SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC - SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
EMBL: U58133; AAC02241.1; -.

```

```

DR HSSP; P14687; 1AMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR003880; Pantane_attach.
DR Pfam; PF00501; AMP-binding; 1.
DR Pfam; PF00550; pp-binding; 1.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; FALSE_NEG.
DR PROSITE; PS00455; AMP_BINDING; 1.
DR PROSITE; PS50075; ACP_DOMAIN; 1.
FW Lysine biosynthesis; Oxidoreductase; NADP; Phosphopantetheine.
KT DOMAIN 852 902 ACRYL CARRIER (ACP). (POTENTIAL).
FT BINDING 884 884 PHOSPHOPANTHETHEINE (POTENTIAL).
SQ SEQUENCE 1391 AA; 154740 MW; B7E641866235BCDD CRC64;

Query Match
Best Local Similarity 92.3%; Score 36; DB 1; Length 1391;
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DDXSXFXGXXQ 12
Db 221 DLDMSGYRGAIQ 232

RESULT 6
LYS2_SCHPO STANDARD; PRT; 1419 AA.
ID LYS2_SCHPO
AC P40976; Q9P770;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Aminodipate-semialdehyde dehydrogenase (EC 1.2.1.31) (Alpha-
DE aminodipate reductase) (Alpha-Ar).
GN LYS1 OR SPAP765.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=96020159; PubMed=8590464;
RA Ford R.A., Bhattacharjee J.K.;
RT "Molecular properties of the lys1+ gene and the regulation of alpha-
RT aminodipate reductase in Schizosaccharomycetes pombe.";
RL Curt. Genet. 28:131-137(1995).

RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymoprez B.,
RA Wellens J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Filtz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler S., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Bartell B.G., Nurse P.;

```

```

RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC - FUNCTION: CATALYZES THE ACTIVATION OF ALPHA-AMINODIPATE BY
CC AMP-DEPENDENT ADENYLATION AND THE REDUCTION OF ACTIVATED
CC ALPHA-AMINODIPATE BY NADPH.
CC - CATALYTIC ACTIVITY: L-2-aminodipate 6-semialdehyde + NADP(+) +
CC H(2)O = L-2-aminodipate + NADPH.
CC - COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETHEINE
CC (POTENTIAL).
CC - PATHWAY: lysine biosynthesis; sixth step.
CC - SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC - SIMILARITY: CONTAINS 1 ACRYL CARRIER DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U15923; AAC15909.1; -.
DR EMBL; AL353014; CAB88271.1; -.
DR HSSP; P14687; 1AMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR003880; Pantane_attach.
DR Pfam; PF00501; AMP-binding; 1.
DR Pfam; PF00550; pp-binding; 1.
DR PROSITE; PS00455; AMP_BINDING; 1.
DR PROSITE; PS50075; ACP_DOMAIN; 1.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.
DR DOMAIN 885 953 ACRYL CARRIER (ACP). (POTENTIAL).
KT Lysine biosynthesis; Oxidoreductase; NADP; Phosphopantetheine.
FW BINDING 916 916 PHOSPHOPANTHETHEINE (POTENTIAL).
FT BINDING 90 91 MT -> IA (IN REF. 1).
FT CONFLICT 196 196 MISSING (IN REF. 1).
FT CONFLICT 487 487 D -> DG (IN REF. 1).
FT CONFLICT 500 500 R -> G (IN REF. 1).
FT CONFLICT 600 602 MISSING (IN REF. 1).
FT CONFLICT 620 621 AR -> GP (IN REF. 1).
FT CONFLICT 712 712 Y -> S (IN REF. 1).
FT CONFLICT 926 928 LR -> PQ (IN REF. 1).
FT CONFLICT 1203 1205 VVV -> AAA (IN REF. 1).
FT CONFLICT 1225 1228 LVVM -> WSK (IN REF. 1).
FT CONFLICT 1239 1239 P -> A (IN REF. 1).
SQ SEQUENCE 1419 AA; 156893 MW; 09AFBEE09F21A0C CRC64;

Query Match
Best Local Similarity 92.3%; Score 36; DB 1; Length 1419;
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DDXSXFXGXXQ 12
Db 271 DLDMSGYRGAIQ 282

RESULT 7
HMA1_BACSU STANDARD; PRT; 432 AA.
ID HMA1_BACSU
AC O07621;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Heme-based aerotactic transducer hemaT.
GN HEMA1.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Nodack M.A., Terpstra P., Holsappel S., Venema G., Bron S.;

```

RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN=168:  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertolo M.G., Bessieres P., Bolotin A., Borchart S.,  
RA Borris R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Brusch C.V., Caldwell B., Capiano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Enrich S.D., Emerson P.T.,  
RA Ertlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,  
RA Giuseppe G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holtsappel S., Hosono S., Hullio M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasanara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koeltter P., Konigstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
RA Pario V., Pohl F.M., Portetle D., Porwolik S., Prescott A.M.,  
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiuchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Yamakoshi A., Yanaka T., Terpstra P., Tognoni A.,  
RA Totsu V., Uchiyama S., Vandendol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambuit R., Wedler E., Wedler H., Weitzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zimstein E., Yoshikawa H., Zanchin A.;  
RT "the complete genome sequence of the gram-positive bacterium Bacillus  
subtilis.";  
RL Nature 390:249-256(1997).  
RN (3)  
RP FUNCTION.  
RX MEDLINE=20140131; PubMed=10676961;  
RA Hou S., Larsen R.W., Boudko D., Riley C.W., Karatan E., Zimmer M.,  
RA Ordal G.W., Altm M.;  
RT "Myoglobin-like aerotaxis transducers in Archaea and Bacteria.";  
RL Nature 403:540-544(2000).  
RN (4)  
RP RESONANCE RAMAN SPECTROSCOPY, AND SUBUNIT.  
RX MEDLINE=21950695; PubMed=11821422;  
RA Aono S., Kato T., Matsuki M., Nakajima H., Ohta T., Uchida T.,  
RA Kitagawa T.;  
RT "Resonance Raman and ligand binding studies of the oxygen-sensing  
RT signal transducer protein Hemat from Bacillus subtilis.";  
RL J. Biol. Chem. 277:13528-13538(2002).  
CC -1- FUNCTION: Heme-containing signal transducer responsible for  
CC aerotaxis, the migratory response toward or away from oxygen.  
CC -1- SUBUNIT: Homotetramer (Probable).  
CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC DR EMBL: Y14084; CAA74545.1; -;  
CC DR EMBL: Z99109; CAB12878.1; -;  
CC DR Subtilist; BG13066; hemat.  
CC InterPro: IPR004089; Chmtaxis\_transd.  
CC Pfam: PF00015; MCPsignal.1.  
CC SMART: SM00283; MA.1.  
CC Transducer; Heme; Complete proteome.  
KW SEQUENCE 432 AA; 48767 MW; 87D065597407DB64 CRC64;

Query Match 89.7%; Score 35; DB 1; Length 432;  
Best Local Similarity 41.7%; Pred No. 65;  
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 12  
|:|:|:|:|:  
Db 399 DVDLQAFGLGLQ 410

RESULT 8  
CKX1.ORYSA STANDARD; PRT; 532 AA.  
ID CKX1.ORYSA  
AC 091DE6;  
AD 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DE Probable cytokinin oxidase precursor (EC 1.4.3.18) (CKO).  
GN P0512G09.9.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Eriatridae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Yamamoto K.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Catalyzes the oxidation of cytokinins, a family of  
CC N(6)-substituted adenine derivatives that are plant hormones,  
CC where the substituent is an isopentenyl group.  
CC -1- CATALYTIC ACTIVITY: N6-(3-methylbut-2-enyl)adenine + H(2)O + O(2)  
CC = adenine + 3-methylbut-2-enal + H(2)O(2).  
CC -1- COFACTOR: FAD (By similarity).  
CC -1- SUBCELLULAR LOCATION: Extracellular (By similarity).  
CC -1- SIMILARITY: Belongs to the oxygen-dependent FAD-linked  
CC oxidoreductase family.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC DR EMBL: AP002836; BAB07927.1; -;  
CC DR EMBL: AP002816; BAB03420.1; -;  
CC DR InterPro: IPR001575; Oxid\_FAD\_Bind.  
CC Pfam: PF01565; FAD\_binding\_4; 1.  
CC DR PROSITE: PS00862; OX2\_COVAL\_FAD; 1.  
CC KW Oxidoreductase; Flavoprotein; FAD; Signal; Glycoprotein.  
CC FT SIGNAL 1 17  
CC FT CHAIN 18 532  
CC FT BINDING 105 105  
CC FT CARBOHYD 52 52  
CC FT CARBOHYD 63 63  
CC FT CARBOHYD 133 133  
CC FT CARBOHYD 321 321  
CC FT CARBOHYD 432 432  
CC FT CARBOHYD 432 432  
CC SEQUENCE 532 AA; 56034 MW; B849D646E3BBA7E CRC64;

Query Match 89.7%; Score 35; DB 1; Length 532;  
Best Local Similarity 41.7%; Pred No. 83;  
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 1 DDXSXFXGXQ 12  
|:|:|:|:|:  
Db 406 DFDGIVFKILQ 417

RESULT 9

CKX1\_MAIZE STANDARD; PRT; 534 AA.  
 ID CKX1\_MAIZE 081158;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cytochrome oxidase 1 precursor (EC 1.4.3.18) (CKO 1).  
 GN CKX1.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 OC Panicoideae; Andropogoneae; Zea.  
 NCBI\_TaxID=4577;  
 [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 417-435; 490-517 AND 524-534.  
 RC STRAIN=cv. Nobilis; TISSUE=kernel;  
 RX MEDLINE=9924666; PubMed=10230061;  
 RA Houba-Herlin N., Peché C., D'Alayer J., Laloue M.;  
 RT "Cytochrome oxidase from Zea mays: purification, cDNA cloning and  
 RL expression in moss protoplasts.";  
 RL Plant J. 17:615-626(1999).  
 [2]  
 RP SEQUENCE FROM N.A., SEQUENCE OF 286-308; 369-377; 388-392 AND 417-431,  
 AND MASS SPECTROMETRY.  
 RX MEDLINE=99160863; PubMed=10049708;  
 RA Morris R.O., Bilyeu K.D., Laskey J.G., Cheikh N.N.;  
 RT "Isolation of a gene encoding a glycosylated cytochrome oxidase from  
 RL maize.";  
 RL Biochem. Biophys. Res. Commun. 255:328-333(1999).  
 [3]  
 RP CHARACTERIZATION, AND TISSUE SPECIFICITY.  
 RX MEDLINE=21068113; PubMed=1154345;  
 RA Bilyeu K.D., Cole J.L., Laskey J.G., Riekhof W.R., Esparza T.J.,  
 RT Kramer M.D., Morris R.O.;  
 RL "Molecular and biochemical characterization of a cytochrome oxidase  
 RT from maize.";  
 RL Plant Physiol. 125:378-386(2001).  
 [4]  
 CC -1- FUNCTION: Catalyzes the oxidation of cytochromes, a family of N(6)-  
 CC substituted adenine derivatives that are plant hormones, where the  
 CC substituent is an isopentenyl group. Cleaves zeatin,  
 CC isopentenyladenine, isopentenyladenosine, zeatin riboside and cis-  
 CC zeatin, but not dihydrozeatin, kinetin and benzylaminopurine.  
 CC -1- CATALYTIC ACTIVITY: N6-(3-methylbut-2-enyl)adenine + H(2)O + O(2)  
 CC => adenine + 3-methylbut-2-enal + H(2)O(2).  
 CC -1- COFACTOR: FAD.  
 CC -1- ENZYME REGULATION: Competitive inhibition by phenylureas.  
 CC -1- SUBUNIT: Monomer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: Expressed in immature kernels and unpollinated  
 CC cobs. Weakly expressed in kernels harvested two weeks after  
 CC anthesis.  
 CC -1- PTM: Glycosylated, with approximately 10 hexose residues per site.  
 CC -1- SIMILARITY: Belongs to the oxygen-dependent FAD-linked  
 CC oxidoreductase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: Y18377; CAA77151.1;  
 CC EMBL: AF044603; AAC27500.1;  
 DR MaizeDB; 194080;  
 DR InterPro; IPR001575; Oxid\_FAD\_bind.  
 DR Pfam; PF01565; FAD\_binding\_4; 1.  
 DR PROSITE; PS00862; OX2\_COVAL\_FAD; 1.  
 KW Oxidoreductase; Flavoprotein; FAD; Signal; Glycoprotein.  
 FT SIGNAL 1 18  
 FT CHAIN 19 534 POTENTIAL  
 FT DOMAIN 339 545 CYTOCHROME OXIDASE 1.  
 FT POLY-ALA.

FT BINDING 105 105 FAD (COVALENT) (PROBABLE).  
 FT CARBOHYD 52 52 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 338 338 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 79 79 A -> G (IN REF. 2).  
 FT CONFLICT 168 168 T -> N (IN REF. 2).  
 FT CONFLICT 254 254 L -> F (IN REF. 2).  
 SQ SEQUENCE 534 AA; 57195 MW; 3D6E71BC4F40918B CRC64;  
 Query Match 89.7%; Score 35; DB 1; Length 534;  
 Best Local Similarity 41.7%; Pred. No. 83;  
 Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 DDXSXFXGXQXQ 12  
 Db 409 DFDGVCFKGILQ 420  
 RESULT 10  
 MANA\_RHOMR STANDARD; PRT; 1021 AA.  
 ID MANA\_RHOMR 49425;  
 AC P49425;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Mannan endo-1,4-beta-mannosidase (EC 3.2.1.78).  
 GN MANA.  
 OS Rhodothermus marinus (Rhodothermus obamensis).  
 OC Bacteria; Bacteroidetes; Sphingobacteriales; Sphingobacteriales;  
 OC Crenothricaceae; Rhodothermus.  
 NCBI\_TaxID=29549;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 43812;  
 RA Pollitz O., Krah M., Borriess R.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic  
 CC linkages in mannans, galactomannans, glucamannans, and  
 CC galactoglucomannans.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 26 OF GLYCOSYL HYDROLASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X90947; CAA62442.1;  
 DR InterPro; IPR000805; Glyco\_hydro\_26.  
 DR Pfam; PF02156; Glyco\_hydro\_26; 1.  
 DR PRINTS; PR00739; GLHYDRLASE26.  
 KW Hydrolase; Glycosidase.  
 SQ SEQUENCE 1021 AA; 115791 MW; A6F7B63109E0F715 CRC64;  
 Query Match 89.7%; Score 35; DB 1; Length 1021;  
 Best Local Similarity 41.7%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DDXSXFXGXQXQ 12  
 Db 363 EADRSVFRGLAQ 374  
 RESULT 11  
 ELHB\_APLCA STANDARD; PRT; 54 AA.  
 ID ELHB\_APLCA 11924;  
 AC P11924;

```

DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Calfin B.
OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidae;
OC Aplysiidae; Aplysia.
OX NCBI_TaxId=6500;
RN [1]
RP SEQUENCE.
RC TISSUE=atrial gland;
RX MEDLINE=86111826; PubMed=3753705;
RA Rothman B.S., Hawke D.H., Brown R.O., Lee T.D., Dehghan A.A.,
RA Shively J.E., Mayerl E.;
RT "Isolation and primary structure of the calfins, three biologically
RT active egg-laying hormone-like peptides from the atrial gland of
RT Aplysia californica."
RL J. Biol. Chem. 261:1616-1623(1986).
RN [2]
RP SEQUENCE.
RC TISSUE=atrial gland;
RX MEDLINE=86224083; PubMed=3711112;
RA Nagle G.T., Painter S.D., Blankenship J.E., Dixon J.D., Kurosky A.;
RT "Evidence for the expression of three genes encoding homologous
RT atrial gland peptides that cause egg laying in Aplysia."
RL J. Biol. Chem. 261:7853-7859(1986).
CC -1- FUNCTION: INJECTED IN SEXUALLY MATURE ANIMALS CALIFIN B EXCITES
CC LB AND LC CELLS OF THE ABDOMINAL GANGLION AND CAUSE EGG-LAYING.
CC -1- SUBUNIT: THIS PROTEIN CONSISTS OF A LARGE 36-RESIDUE SUBUNIT,
CC BOUND BY A SINGLE DISULFIDE-BOND TO A SMALL 18-RESIDUE SUBUNIT.
CC -1- MISCELLANEOUS: CALIFIN B PROBABLY DERIVES FROM POLYPEPTIDE B,
CC WHICH IS ALSO THE PRECURSOR FOR PEPTIDE B.
CC -1- SIMILARITY: TO ELH/ACIDIC PEPTIDE.
DR PIR; B26295; B26295.
DR PIR; D26306; D26306.
KW Hormone; Amidation; Neuropeptide.
FT CHAIN 1 36 LARGE SUBUNIT.
FT NON_CONS 36 37
FT CHAIN 37 54 SMALL SUBUNIT.
FT DISULFID 25 53
FT MOD_RES 36 36
SQ SEQUENCE 54 AA; 6057 MW; C6B60521BC900914 CRC64;

Query Match 87.2%; Score 34; DB 1; Length 54;
Best Local Similarity 45.5%; Pred. No. 9;
Matches 5; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 DDXSXFXGXX 11
ID ELHC_APLCA STANDARD; PRT; 54 AA.
AC p11925;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Calfin C.
OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidae;
OC Aplysiidae; Aplysia.
OX NCBI_TaxId=6500;
RN [1]
RP SEQUENCE.
RC TISSUE=atrial gland;
RX MEDLINE=86111826; PubMed=3753705;
RA Rothman B.S., Hawke D.H., Brown R.O., Lee T.D., Dehghan A.A.,
RA Shively J.E., Mayerl E.;
RT "Isolation and primary structure of the calfins, three biologically
RT active egg-laying hormone-like peptides from the atrial gland of
RT Aplysia californica."

```

```

RL J. Biol. Chem. 261:1616-1623(1986).
RN [2]
RP SEQUENCE.
RC TISSUE=atrial gland;
RX MEDLINE=86224083; PubMed=3711112;
RA Nagle G.T., Painter S.D., Blankenship J.E., Dixon J.D., Kurosky A.;
RT "Evidence for the expression of three genes encoding homologous
RT atrial gland peptides that cause egg laying in Aplysia."
RL J. Biol. Chem. 261:7853-7859(1986).
CC -1- FUNCTION: INJECTED IN SEXUALLY MATURE ANIMALS CALIFIN C EXCITES
CC LB AND LC CELLS OF THE ABDOMINAL GANGLION AND CAUSE EGG-LAYING.
CC -1- SUBUNIT: THIS PROTEIN CONSISTS OF A LARGE 36-RESIDUE SUBUNIT,
CC BOUND BY A SINGLE DISULFIDE-BOND TO A SMALL 18-RESIDUE SUBUNIT.
CC -1- SIMILARITY: TO ELH/ACIDIC PEPTIDE.
DR PIR; C26295; C26295.
DR PIR; C26306; C26306.
KW Hormone; Neuropeptide; Amidation.
FT CHAIN 1 36 LARGE SUBUNIT.
FT NON_CONS 36 37
FT CHAIN 37 54 SMALL SUBUNIT.
FT MOD_RES 36 36
FT DISULFID 25 53
SQ SEQUENCE 54 AA; 6085 MW; EB99D521BC900917 CRC64;

Query Match 87.2%; Score 34; DB 1; Length 54;
Best Local Similarity 45.5%; Pred. No. 9;
Matches 5; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 DDXSXFXGXX 11
ID Y843_AQUAE STANDARD; PRT; 133 AA.
AC 067011;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_843.
GN AQ_843.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxId=63363;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Shead M.A., Keller M., Anjlay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
CC -1- SIMILARITY: BELONGS TO THE UPF0079 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000710; AAC06981.1; -.
DR InterPro; IPR003442; UPF0079.
DR Pfam; PF02367; UPF0079; 1.
DR TIGRPFAMs; TIGR00150; UPF0079; 1.
KW Hypothetical protein; ATP-binding; Complete proteome.
NP_BIND 36 43 ATP (POTENTIAL).
SQ SEQUENCE 133 AA; 15246 MW; 45C990DEDD12C4138 CRC64;

```



Query Match 87.2%; Score 34; DB 1; Length 133;  
 Best Local Similarity 45.5%; Pred No. 26;  
 Matches 5; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDXSXFSGXX 11  
 DB 89 DFDYSEFIBG 99

RESULT 14  
 ID ELHA\_APLCA STANDARD; PRT; 173 AA.  
 AC P01360; P11923;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-AUG-1990 (Rel. 15, Last annotation update)  
 DE Atrial gland peptide A precursor (ELH-18) [contains: Callicin A].  
 OS Aplysia californica (California sea hare).  
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidea;  
 OC Aplysiidae; Aplysia.  
 OX NCBI\_TaxID=6500;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85263781; PubMed=4020422;  
 RA Mahon A.C., Nambu J.R., Taussig R., Shyamala M., Roach A.,  
 RA Scheller R.H.;  
 RT "Structure and expression of the egg-laying hormone gene family in  
 RT Aplysia";  
 RL J. Neurosci. 5:1872-1880(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83129416; PubMed=6687446;  
 RA Scheller R.H., Jackson J.F., McAllister L.B., Rothman B.S.,  
 RA Mayeri E., Axel R.;  
 RT "A single gene encodes multiple neuropeptides mediating a stereotyped  
 RT behavior";  
 RL Cell 32:7-22(1983).  
 RN [3]  
 RP SEQUENCE OF 36-69 (PEPTIDE A).  
 RX MEDLINE=80190239; PubMed=6929554;  
 RA Heller E., Kaczmarek L.K., Hunkapiller M.W., Hood L.E.,  
 RA Strumwasser F.;  
 RT "Purification and primary structure of two neuroactive peptides that  
 RT cause bag cell afterdischarge and egg-laying in Aplysia";  
 RL Proc. Natl. Acad. Sci. U.S.A. 77:2328-2332(1980).  
 RN [4]  
 RP SEQUENCE OF 117-173 (CALICIN A).  
 RC TISSUE=Atrial gland;  
 RX MEDLINE=86111826; PubMed=3753705;  
 RA Rothman B.S., Hawke D.H., Brown R.O., Lee T.D., Dehghan A.A.,  
 RA Shively J.E., Mayeri E.;  
 RT "Isolation and primary structure of the callicins, three biologically  
 RT active egg-laying hormone-like peptides from the atrial gland of  
 RT Aplysia californica";  
 RL J. Biol. Chem. 261:1616-1623(1986).  
 RN [5]  
 RP SEQUENCE OF 22-34; 131-152 AND 156-173.  
 RX MEDLINE=88243802; PubMed=3379066;  
 RA Nagle G.T., Painter S.D., Blankenship J.E., Kurosky A.;  
 RT "Proteolytic processing of egg-laying hormone-related precursors in  
 RT Aplysia. Identification of peptide regions critical for biological  
 RT activity";  
 RL J. Biol. Chem. 263:9223-9237(1988).  
 CC -1- FUNCTION: THE ATRIAL GLAND PEPTIDE A & PEPTIDE B PRECURSORS ARE  
 CC THE SOURCE OF THE 2 PEPTIDES THAT, UPON RELEASE FROM THIS  
 CC REPRODUCTIVE SYSTEM GLAND, INITIATE THE EGG-LAYING PROCESS BY  
 CC EXITING THE BAG CELL NEURONS. THESE NEURONS, CLUSTERED IN NEURAL  
 CC CONNECTIVES NEAR THE ABDOMINAL GANGLION, IN TURN RELEASE OTHER  
 CC PEPTIDES THAT ACT DIRECTLY ON THE GANGLION & ALSO, VIA THE  
 CC CIRCULATING HEMOLYPH, ON MANY OTHER ORGANS TO CONTROL THE  
 CC PHYSIOLOGICAL PROCESSES OF EGG-LAYING. ONE OF THESE OTHER PEPTIDES  
 CC IS THE EGG-LAYING HORMONE.

CC -1- FUNCTION: INJECTED IN SEXUALLY MATURE CALIFIN A EXCITES  
 CC LB AND LC CELLS OF THE ABDOMINAL GANGLION AND CAUSE EGG-LAYING.  
 CC -1- SUBUNIT: CALIFIN A CONSISTS OF A 36-RESIDUE LARGE SUBUNIT BOUND  
 CC BY A SINGLE DISULFIDE BOND TO A 18-RESIDUE SMALL SUBUNIT.  
 CC -1- SIMILARITY: TO ELH PRECURSOR AND ATRIAL GLAND PEPTIDE B PRECURSOR.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>  
 CC or send an email to [license@isb-stb.ch](mailto:license@isb-stb.ch)).  
 CC -----  
 CC EMBL: M29350; AAA27750.1; -  
 CC EMBL: J01017; AAA27742.1; -  
 CC DR PIR: A01630; GOGANA.  
 CC DR PIR: A26295; A26295.  
 CC DR PIR: B28108; B28108.  
 CC DR InterPro: IPR003424; ELH.  
 CC Pfam: PF02323; ELH: 1.  
 CC Cleavage on pair of basic residues; Neuropeptide; Signal; Amidation.  
 CC KW SIGNAL  
 CC FT SIGNAL 1 21 PROBABLE.  
 CC FT PEPTIDE 36 69 ATRIAL GLAND PEPTIDE A.  
 CC FT CHAIN 117 173 CALIFIN A.  
 CC FT PEPTIDE 117 152 CALIFIN A LARGE SUBUNIT.  
 CC FT PEPTIDE 156 173 CALIFIN A SMALL SUBUNIT.  
 CC FT DISULFID 141 172  
 CC FT MOD\_RES 69 69  
 CC FT MOD\_RES 152 152  
 CC FT CONFLICT 172 173  
 CC FT SEQUENCE 173 AA; 19283 MW; E71D0A340427B0B6 CRC64;  
 CS -> POLKTSNULD (IN REF. 2).  
 CS -> PROVIDE AMIDE GROUP).

Query Match 87.2%; Score 34; DB 1; Length 173;  
 Best Local Similarity 45.5%; Pred. No. 35;  
 Matches 5; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDXSXFSGXX 11  
 DB 156 DSDVSLFNGDL 166

RESULT 15  
 ID FLIN\_AGR5 STANDARD; PRT; 179 AA.  
 AC 057259;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Flagellar motor switch protein flin.  
 GN FLIN OR FLIN OR ATU0562 OR AGR\_C.988.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=176299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97305159; PubMed=9161425;  
 RA Deakin W.J., Parker V.E., Loake G.J., Shaw C.H.;  
 RT "The Agrobacterium tumefaciens motor gene, motA, is in a linked  
 RT cluster with the flagellar switch protein genes, flig, flin and  
 RT flin";  
 RL Gene 189:139-141(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608550; PubMed=11743193;  
 RA Wood D.W., Secubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
 RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyee D., Sr.,  
 RA Chapman P., Glendinning J., Deatherage G., Gillet W., Grant C.,  
 RA Kutayav T., Levy R., Li M.-J., McClelland E., Palmeri A.,  
 RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,  
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,



```

RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nestler E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE-21608551; PubMed-11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Ouello B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmel K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursun J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
CC -!- FUNCTION: FLIN IS ONE OF THREE PROTEINS (FLIG, FLIN, FLIM) THAT
CC FORM A SWITCH COMPLEX THAT IS PROPOSED TO BE LOCATED AT THE BASE
CC OF THE BASAL BODY. THIS COMPLEX INTERACTS WITH THE CHEY AND CHEZ
CC CHEMOTAXIS PROTEINS, IN ADDITION TO CONTRACTING COMPONENTS OF THE
CC MOTOR THAT DETERMINE THE DIRECTION OF FLAGELLAR ROTATION (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (Potential).
CC -!- SIMILARITY: BELONGS TO THE FLIN/MOPA/SPAO FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL; X65584; CAA46543.1; -
DR EMBL; U63290; AAC45323.1; -
DR EMBL; U95165; AAB71780.1; -
DR EMBL; AE009025; AAI41579.1; -
DR EMBL; AE007990; AAK86374.1; ALT_INT.
DR InterPro: IPR001172; Flagellar_FLIN.
DR Pfam: PF01052; SpOA.1.
DR PRINTS: PR00956; FLGMOTORFLIN.
KW Chemotaxis; Flagella; Flagellar rotation; Inner membrane;
KW Complete proteome.
SQ SEQUENCE 179 AA; 18803 MW; 5D441C71B380D432 CRC64;

```

Query Match 87.2%; Score 34; DB 1; Length 179;  
Best Local Similarity 45.5%; Pred. No. 37;  
Matches 5; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDXSXFXGXX 11  
|:|:|:|:|:  
Db 60 DNDLSAFGGGA 70

Search completed: March 18, 2003, 08:30:15  
Job time : 8.5 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.P5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 08:28:17 ; Search time 26 Seconds  
(Without alignments)  
95.099 Million cell updates/sec

Title: US-09-812-485A-43  
Perfect score: 39  
Sequence: 1 DXDXSFXGXGXQ 12

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	100.0	272	16 0990K6	0990K6 streptomyc
2	39	100.0	500	6 095XG5	095XG5 macaca fasc
3	39	100.0	525	4 09NQ76	09NQ76 homo sapien
4	39	100.0	555	6 09GM13	09GM13 macaca fasc
5	39	100.0	555	6 09N076	09N076 macaca fasc
6	39	100.0	1832	16 092J73	092J73 rhizobium m
7	36	92.3	94	5 095P90	095P90 mesobutius
8	36	92.3	137	10 094K15	094K15 tritlicum ae
9	36	92.3	137	10 094K14	094K14 hordeum vul
10	36	92.3	251	3 013813	013813 schizosacch
11	36	92.3	285	5 08T8Q9	08T8Q9 drosophila
12	36	92.3	308	5 095RB4	095RB4 drosophila
13	36	92.3	328	5 08SUP6	08SUP6 encephalito
14	36	92.3	328	5 09VUT8	09VUT8 encephalito
15	36	92.3	358	10 095J35	095J35 arabidopsis
16	36	92.3	430	10 08RX19	08RX19 arabidopsis

17	36	92.3	578	2 08RJP2	08RJP2 erwania chr
18	36	92.3	649	10 081620	081620 arabidopsis
19	36	92.3	665	10 095N59	095N59 arabidopsis
20	36	92.3	756	5 08WSK6	08WSK6 dermacentor
21	36	92.3	778	5 09U329	09U329 drosophila
22	36	92.3	778	5 09VP07	09VP07 drosophila
23	36	92.3	1133	10 004626	004626 arabidopsis
24	36	92.3	1174	3 09P307	09P307 neurospora
25	35	89.7	138	17 097M30	097M30 sulfolobus
26	35	89.7	200	2 09L8E6	09L8E6 vibrio harv
27	35	89.7	226	16 098L23	098L23 rhizobium l
28	35	89.7	257	16 09PDK6	09PDK6 xylella fas
29	35	89.7	450	16 09HTH1	09HTH1 pseudomonas
30	35	89.7	500	5 002050	002050 caenorhabdi
31	35	89.7	529	5 091862	091862 caenorhabdi
32	35	89.7	611	13 090646	090646 gallus gall
33	35	89.7	617	13 090647	090647 gallus gall
34	35	89.7	854	10 042526	042526 arabidopsis
35	35	89.7	858	10 023647	023647 arabidopsis
36	35	89.7	1009	10 09M2T8	09M2T8 arabidopsis
37	35	89.7	1199	5 018782	018782 caenorhabdi
38	35	89.7	1305	5 09W1R4	09W1R4 drosophila
39	35	89.7	2556	11 091XW2	091XW2 mus musculu
40	35	89.7	3940	3 09HE19	09HE19 neurospora
41	34	87.2	109	2 08RPH9	08RPH9 desulfitoba
42	34	87.2	120	2 049573	049573 mycobacteri
43	34	87.2	120	16 09CCA4	09CCA4 mycobacteri
44	34	87.2	131	13 09GW20	09GW20 hippoglossu
45	34	87.2	131	16 033220	033220 mycobacteri

## ALIGNMENTS

### RESULT 1

ID 0990K6 PRELIMINARY; PRT; 272 AA.

AC 0990K6; 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Hypothetical protein SCPl.337C.  
GN SCPl.337C AND SCPl.17.  
OS Streptomyces coelicolor.  
OG Plasmid SCPl.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinowitsch E., Rajandream W.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomyete Streptomyces  
coelicolor A3(2)";  
RL Nature 417:141-147(2002).  
DR EMBL: AL590464; CAC38682.1; -;  
DR EMBL: AL590463; CAC36539.1; -;  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 272 AA; 29985 MW; 2D7D6D25B936E9E0 CRC64;

Query Match 100.0%; Score 39; DB 16; Length 272;  
Best Local Similarity 50.0%; Pred. No. 30;  
Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DXDXSFXGXGXQ 12

DB 18 DEDSDSEFWFAQ 29

## RESULT 2

ID 095KG5 PRELIMINARY; PRT; 500 AA.

AC 095KG5;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Hypothetical 55.6 kDa protein.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Macaca.  
 OX NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TEMPORAL LOBE RIGHT;  
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,  
 RA Suzuki Y., Sugano S., Hashimoto K.;  
 RT "Isolation of full-length cDNA clones from macaque brain cDNA  
 libraries.";  
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB060891; BAB46894.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 500 AA; 55577 MW; 918D265ADBEDC7BC CRC64;

## Query Match

Best Local Similarity 100.0%; Score 39; DB 6; Length 500;  
 Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDXSXFXGXQ 12  
 DB 225 DNDMSPFSGDQ 236

## RESULT 3

ID 09N076 PRELIMINARY; PRT; 525 AA.

AC 09N076;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Matrix extracellular phosphoglycoprotein precursor.  
 GN MEPE.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BONE;  
 RA MEDLINE=20399567; PubMed=10945470;  
 RA Rowe P.S.N., De Zoysa P.A., Dong R., Wang H.R., White K.E.,  
 RA Econs M.J., Oudet C.L.;  
 RT "MEPE, a new gene expressed in bone marrow and tumors causing  
 RT osteomalacia.";  
 RL Genomics 67:54-68(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=21309068; PubMed=11414762;  
 RA Argiro L., Desparats M., Glorieux F.H., Ecartot B.;  
 RT "Mepe, the gene encoding a tumor-secreted protein in oncogenic  
 RT hypophosphatemic osteomalacia, is expressed in bone.";  
 RL Genomics 74:342-351(2001).  
 DR EMBL; AJ276396; CAB97250.1; -;  
 KW EMBL; AF325916; AAK70343.1; -;  
 RN Signal.  
 FT SIGNAL 1 17 POTENTIAL.  
 FT CHAIN 18 525 MATRIX EXTRACELLULAR PHOSPHOGLYCOPROTEIN.  
 SQ SEQUENCE 525 AA; 58419 MW; 0977CA6E871CA95 CRC64;

Query Match 100.0%; Score 39; DB 4; Length 525;  
 Best Local Similarity 50.0%; Pred. No. 68;  
 Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDXSXFXGXQ 12  
 DB 249 DNDISPFSGDQ 260

## RESULT 4

ID 09GM13 PRELIMINARY; PRT; 555 AA.

AC 09GM13;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE MEPE protein.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Macaca.  
 OX NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN PARIENTAL LOBE;  
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,  
 RA Suzuki Y., Sugano S., Hashimoto K.;  
 RT "Isolation of full-length cDNA clones from macaque brain cDNA  
 libraries.";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB050259; BAB17010.1; -;  
 KW SEQUENCE 555 AA; 61950 MW; BD4D47E88038A9E2 CRC64;

## Query Match

Best Local Similarity 100.0%; Score 39; DB 6; Length 555;  
 Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDXSXFXGXQ 12  
 DB 280 DNDMSPFSGDQ 291

## RESULT 5

ID 09N076 PRELIMINARY; PRT; 555 AA.

AC 09N076;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE Matrix extracellular phosphoglycoprotein.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Macaca.  
 OX NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,  
 RA Suzuki Y., Sugano S., Hashimoto K.;  
 RT "Isolation of full-length cDNA clones from macaque brain cDNA  
 libraries.";  
 RL Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB046056; BAB01638.1; -;  
 KW SEQUENCE 555 AA; 61979 MW; CCE1A98ADA19E4 CRC64;

## Query Match

Best Local Similarity 100.0%; Score 39; DB 6; Length 555;  
 Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDXSXFXGXQ 12  
 DB 280 DNDMSPFSGDQ 291

## RESULT 6

092JUT3 PRELIMINARY; PRT; 1832 AA.  
ID 092JUT3  
AC 092JUT3;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)  
DE 01-MAR-2002 (TREMBlrel. 20, last annotation update)  
OS Hypothetical protein R03146.  
GN R03146 OR SMC03761.  
OC Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=21396507; PubMed=11481430;  
RA Capela D., Barloy-Hubier F., Gouzy J., Bothe G., Ampe F., Batut J.,  
RA Boizard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,  
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,  
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;  
RT "Analysis of the chromosome sequence of the legume symbiont  
RT Sinorhizobium meliloti strain 1021."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
DR EMBL: AL591793; CAC47725.1; -.  
DR InterPro: IPR001410; DEAD.  
DR InterPro: IPR001650; Helicase\_C.  
DR Pfam: PF00270; DEAD; 1.  
DR Pfam: PF00271; helicase\_C; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 1832 AA; 202209 MW; 23053ABF8833BB01 CRC64;

## Query Match

Best Local Similarity 100.0%; Score 39; DB 16; Length 1832;  
Best Local Similarity 50.0%; Pred. No. 3; 2e+02;  
Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 DXDXSFXGXGXQ 12  
|:|:|:|:|:|:|

DB 1627 DDPSEFRGGRQ 1638

## RESULT 7

095P90 PRELIMINARY; PRT; 94 AA.  
ID 095P90  
AC 095P90;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)  
DE 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
OS Putative toxin-like peptide KTXIP2.  
OS Mesodutucha martensii (Manchurian scorpion) (Butus martensii).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;  
OC Butoidea; Butidae; Butus.  
OX NCBI\_TaxID=34649;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=VENOM GLAND.  
RA Zhu S., Li W.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF155368; AAK61818.1; -.  
SQ SEQUENCE 94 AA; 10549 MW; F152AE78BE2FF7E0 CRC64;

## Query Match

Best Local Similarity 92.3%; Score 36; DB 5; Length 94;  
Best Local Similarity 41.7%; Pred. No. 34;  
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1 DXDXSFXGXGXQ 12  
|:|:|:|:|:|:|

DB 75 DYDNSFFFGALE 86

## RESULT 8

094K15 PRELIMINARY; PRT; 137 AA.  
ID 094K15  
AC 094K15;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)  
DE 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
OS Cytochrome dehydrogenase (Fragment).  
OS Triticum aestivum (Wheat).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
OC Triticeae; Triticum.  
OX NCBI\_TaxID=4565;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Galuszka P., Frebort I., Sebel M., Yamada M., Pec P.;  
RT "Cloning of the genes encoding cytochrome dehydrogenase in wheat and  
RT barley."  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF362471; AAK51494.1; -.  
FT NON\_TER 1  
FT NON\_TER 137  
SQ SEQUENCE 137 AA; 15553 MW; 37510BCA1277EAE7 CRC64;

## Query Match

Best Local Similarity 92.3%; Score 36; DB 10; Length 137;  
Best Local Similarity 41.7%; Pred. No. 54;  
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1 DXDXSFXGXGXQ 12  
|:|:|:|:|:|:|

DB 16 DFDRAVFRGILQ 27

## RESULT 9

094K14 PRELIMINARY; PRT; 137 AA.  
ID 094K14  
AC 094K14;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)  
DE 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
OS Cytochrome dehydrogenase (Fragment).  
OS Hordeum vulgare (Barley).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
OC Triticeae; Hordeum.  
OX NCBI\_TaxID=4513;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Galuszka P., Frebort I., Sebel M., Yamada M., Pec P.;  
RT "Cloning of the genes encoding cytochrome dehydrogenase in wheat and  
RT barley."  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF362472; AAK51495.1; -.  
FT NON\_TER 1  
FT NON\_TER 137  
SQ SEQUENCE 137 AA; 15550 MW; 675D4E5610B4B682 CRC64;

## Query Match

Best Local Similarity 92.3%; Score 36; DB 10; Length 137;  
Best Local Similarity 41.7%; Pred. No. 54;  
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1 DXDXSFXGXGXQ 12  
|:|:~:~:~:~:~:~:~

DB 16 DFDRAVFRGILQ 27

## RESULT 10

013813 PRELIMINARY; PRT; 144 AA.  
ID 013813  
AC 013813;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)

DE Putative mitochondrial import receptor subunit (Mitochondrial 22 kDa  
DE outer membrane protein) (Translocase of outer membrane 22 kDa subunit)  
DE (MAS17 protein) (Mitochondrial 17 kDa assembly protein).  
GN SPAC17H9.16.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.,  
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: CENTRAL COMPONENT OF THE RECEPTOR COMPLEX RESPONSIBLE  
CC FOR THE RECOGNITION AND TRANSLLOCATION OF CYTOSOLICALLY SYNTHESIZED  
CC MITOCHONDRIAL PREPROTEINS (BY SIMILARITY).  
CC -1- SUBUNIT: FORMS PART OF THE RECEPTOR COMPLEX THAT CONSISTS OF AT  
CC LEAST 8 DIFFERENT PROTEINS (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MITOCHONDRIAL  
CC OUTER MEMBRANE (BY SIMILARITY).  
CC -1- SIMILARITY: TO YEAST TOM22.  
DR EMBL: Z98597; CAB1225.1; -  
KW TIGRFRAMS: TIGR00986; 3a0801s05com22.1.  
KM Hypothetical protein; Transport; Protein transport; Outer membrane;  
KM Mitochondrion; Transmembrane.  
FT DOMAIN 1 90 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 91 112 POTENTIAL.  
FT DOMAIN 113 144 INTERMEMBRANE (POTENTIAL).  
SQ SEQUENCE 144 AA; 15912 MW; 8AA1823E28C6589E CRC64;

Query Match 92.3%; Score 36; DB 3; Length 144;  
Best Local Similarity 41.7%; Pred. No. 57;  
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDXSXFQXQ 12  
DB 37 DSDSEFEGLEE 48

RESULT 11  
Q8TR09 PRELIMINARY; PRT; 251 AA.  
AC Q8TR09;  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE SP03278P.  
GN ARU.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GB-M1;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champagne M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,  
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,  
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
RA Celisner S.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY075572; AAL68378.1; -  
SQ SEQUENCE 251 AA; 27011 MW; FA95DB8B05146935 CRC64;

Query Match 92.3%; Score 36; DB 5; Length 251;  
Best Local Similarity 41.7%; Pred. No. 11e+02;  
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDXSXFQXQ 12  
DB 194 DDDGAGTGGQO 205

RESULT 12  
Q95RB4 PRELIMINARY; PRT; 285 AA.  
AC Q95RB4;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE LD46333p.  
GN CG9293.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY061507; AAL29055.1; -  
DR FLYBASE: FBgn0032516; CG9293.  
DR InterPro: IPR001965; znf\_PHD.  
DR Pfam: PF00628; PHD; 1.  
SQ SEQUENCE 285 AA; 32080 MW; C098B515673FEAC CRC64;

Query Match 92.3%; Score 36; DB 5; Length 285;  
Best Local Similarity 41.7%; Pred. No. 1.3e+02;  
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDXSXFQXQ 12  
DB 104 DNDLRFEGEIQ 115

RESULT 13  
Q8SUF6 PRELIMINARY; PRT; 308 AA.  
AC Q8SUF6;  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Hypothetical protein EC010\_0650.  
GN EC010\_0650.  
OS Encephalitozoon cuniculi.  
OC Eukaryota; Microsporidia; Unikaryoniidae; Encephalitozoon.  
OC Eukaryota; Microsporidia; Unikaryoniidae; Encephalitozoon.  
OX NCBI\_TaxID=6035;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GB-M1;  
RA Genoscope;  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GB-M1;  
RX MEDLINE=21576510; PubMed=11719806;  
RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,  
RA Prensier G., Barbe V., Peyretallade E., Brotlier P., Wincker P.,  
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gony M.,  
RA Weissenbach J., Vivares C.P.;  
RT "Genome sequence and gene compaction of the eukaryote parasite  
RT Encephalitozoon cuniculi.";  
RL Nature 414:450-453(2001).  
DR EMBL: AL590449; CAD25784.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 308 AA; 34814 MW; F09148E030A0EEC2 CRC64;

Query Match 92.3%; Score 36; DB 5; Length 308;

Best Local Similarity 41.7%; Pred. No. 1.5e+02;  
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDXSXFEXGXQ 12  
|:|:|:|:|:|:  
Db 276 DVLSTYFDGRKE 287

## RESULT 14

Q9VJ78 PRELIMINARY; PRT; 328 AA.  
AC Q9VJ78;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE CG9293 protein.  
GN CG9293.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxId=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;

Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
R Brandon R.C., Rogers J.-H.C., Blazel R.G., Champe M., Pfeiffer B.D.,  
Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,  
Abril J.F., Agbayani A., An H.-J., Andrews-Plamkoc C., Baldwin D.,  
Bailly R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,  
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,  
Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,  
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
Ye J., Yeh R., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003640; AAF53297.1; -;  
DR FlyBase; FBgn0032516; CG9293.  
DR InterPro; IPR001965; znf\_PHD.  
DR Pfam; PF00628; PHD; 1.  
DR SMART; SM00249; PHD; 1.  
SQ SEQUENCE 328 AA; 37051 MW; 0B6BC8F0684891B7 CRC64;

Query Match 92.3%; Score 36; DB 5; Length 328;  
Best Local Similarity 41.7%; Pred. No. 1.6e+02;

Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDXSXFEXGXQ 12  
|:|:|:|:|:|:  
Db 125 DNDLAREGEIQ 136

## RESULT 15

Q9SJ35 PRELIMINARY; PRT; 358 AA.  
AC Q9SJ35;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE At2g05260 protein.  
GN At2g05260.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxId=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=20083487; PubMed=10617197;

Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
Fuji C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,  
Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
Salzberg S.L., Fraser C.M., Venter J.C.;  
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
thaliana.";  
RL Nature 402:761-768(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC007018; AAD29063.1; -;  
DR InterPro; IPR002921; Lipase\_3.  
DR InterPro; IPR002052; N6\_Mtase.  
DR Pfam; PF01764; Lipase\_3; 1.  
DR PROSITE; PS00092; N6\_Mtase; UNKNOWN\_1.  
SQ SEQUENCE 358 AA; 40130 MW; 694F8F5E5B9A6116 CRC64;

Query Match 92.3%; Score 36; DB 10; Length 358;  
Best Local Similarity 41.7%; Pred. No. 1.8e+02;  
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDXSXFEXGXQ 12  
|:|:|:|:|:|:  
Db 74 DADFSIIGCIFE 85

Search completed: March 18, 2003, 08:31:17  
Job time : 29 secs

**THIS PAGE BLANK (USPTO)**